# Implementing Punishment and Reward in the Public Goods Game: The Effect of Individual and Collective Decision Rules*

Nynke van Miltenburg†[1], Vincent Buskens[1,2], Davide Barrera[3,1], Werner Raub[1]

[1]Department of Sociology/ICS, Utrecht University, The Netherlands
[2]Erasmus School of Law, Erasmus University, The Netherlands
[3]Department of Social Sciences, University of Turin, Italy

**Abstract**

Sanctions are effective means for establishing cooperation in social dilemmas. We compare a setting where actors individually decide whom to sanction to a setting where sanctions are only executed when actors collectively agree whom to target. Collective decision rules are problematic due to the difficulty of reaching agreement on sanctions. However, when a decision is made collectively, antisocial sanctioning of individual actors is ruled out. Therefore, sanctions implemented through collective decisions are more likely to be in the interest of the whole group. We employ a laboratory experiment where subjects play one-shot Public Goods Games with opportunities for punishment or reward that can be implemented either by an individual, a majority, or unanimously. For both punishment and reward, contribution levels are higher in the individual than the majority condition, and higher under majority than unanimity. Often, majority agreement or unanimity was not reached.

**Keywords:** Public Goods Game; collective decision rule; conditional cooperation; reward

*Comments of and discussions with colleagues of our Utrecht group "Cooperation in Social and Economic Relations" are gratefully acknowledged. The paper is part of the project "The Feasibility of Cooperation under Various Sanctioning Institutions," funded by the Netherlands Organization for Scientific Research (NWO, grant 400-09-159). Earlier drafts were presented in 2011 at the Symposium of the International Sociological Association (ISA) Research Committee 45 on Rational Choice, at the Fourth Conference of the European Network of Analytical Sociologists (ENAS), and at the Nuffield Workshop on Experimental Research on Social Dilemmas.

†Corresponding author: Department of Sociology/ICS, Utrecht University, Padualaan 14, PO Box 80.140, 3508 TC Utrecht, The Netherlands, T +31 302534813, n.vanmiltenburg@uu.nl

# 1 INTRODUCTION

A public good is characterized by non-excludability: once it is produced, all actors can enjoy its benefits regardless of their contribution to the provision of the good (Olson, 1965). Since public good provision is costly, this implies a tension between individual and collective interest. While mutual cooperation leads to the best possible group outcome, individuals have an incentive to free-ride on the contributions of others.

Contributions to public goods can be supported by sanctions, that is, the opportunity for actors to punish or reward each other. Experimental research established that high contributions are maintained when sanctioning is possible (Fehr and Gächter, 2000; Sefton et al., 2007). However, which method of assigning sanctions works best in enforcing cooperation remains an open question (Gächter and Thöni, 2011). For example, it is unclear whether contributions are higher when the decision of whom to sanction is made individually or when it is made collectively. A sanctioning system with an individual decision rule (IDR) is a system in which every actor individually decides whom to sanction and pays the associated costs. A sanctioning system with a collective decision rule (CDR) is a system in which sanctions are executed only when multiple actors agree and each actor pays the cost of sanctioning.

In real-life public good problems, actors often employ a sanctioning institution with a CDR. For example, Ostrom (1990) and Veszteg and Narhetali (2010) describe small communities where group members successfully enforce collective action through collective sanctioning decisions. Typically, members of the community regularly meet to identify free-riders and decide upon their punishment, for example in a vote. Also, in international cooperation, nations use collective sanctioning decision rules to ensure provision of global public goods such as international security and economic stability. Sanctioning decisions are usually taken by a variant of majority voting. Unanimity voting is uncommon, because it gives every individual nation the opportunity to veto a sanction, thereby making collective organizations ineffective decision makers (http://www.europa.eu, 2010; http://www.un.org, 2010).

So far, there is limited experimental research comparing the effect of sanctioning through IDRs and CDRs in public good problems. Casari and Luini (2009) find that, compared to an IDR, contributions to public goods are higher when punishment is only carried out if at least two out of four actors want to punish the fifth member of their group. Their results, however, leave a number of unresolved issues, in which the current paper provides further insight.

First, it is unclear how the effect of a CDR on contribution depends upon the proportion of actors required to agree for a sanction to be implemented. On the one hand, the higher the proportion required, the less likely it will be that a sufficient number of actors agrees on the necessity of sanctioning and is willing to incur the associated costs (cf. Buchanan and Tullock, 1967). Thus, while under an IDR all desired sanctions are carried out by definition, under CDRs there is a higher chance that free-riders remain unpunished or contributors unrewarded. On the other hand, under an IDR individuals might decide to use sanctions in ways that hurt contribution and thereby result in decreasing payoffs for the group, i.e. to reward free-riders or to punish contributors (Casari and Luini, 2009). Consequently, the more actors collectively agree that a certain group member should be sanctioned, the higher the chance that this sanction will be in the collective interest, that is, in accordance with enforcing contributions to the public good. In the current paper, we address the effect of the required proportion of consenting actors on contribution levels by comparing contributions under an IDR to a CDR for which majority and a CDR for which unanimity is required.

Second, theoretical arguments and empirical results on punishment cannot be straightforwardly generalized to reward. For example, to maintain cooperation a reward has to be allocated every time an actor contributes. Conversely, the mere threat of punishment can be enough to deter free-riding (Dari-Mattiacci and De Geest, 2010). This implies that punishment and rewards differ in efficiency. Empirically, it has been shown that punishment and rewards differ also in terms of efficacy (Sefton et al., 2007; Wiedemann et al., 2012). We therefore study decision rules for assigning both punishment and reward.

Third, predictions regarding the effect of decision rules on contribution depend on assumptions about the preferences of individual actors. For example, on which proportion of actors is willing to sanction, to what extent they sanction in accordance with enforcing cooperation, and how they adjust their contribution level in anticipation of being sanctioned. We therefore summarize existing knowledge on individual behaviour in the PGG with sanctions. Subsequently, we apply this to predict behaviour in the PGG with different decision rules, and with punishment or reward. We thus assess through which mechanisms our empirical extensions could result in different contribution levels between sanctioning systems.

The paper is structured as follows. In the theory section, we review the literature on behaviour in public good problems with opportunities for sanctioning. Subsequently, we develop hypotheses about contribution and sanctioning behaviour, and on how this behaviour of individuals leads to different contribution levels under IDRs and CDRs. Individual-level and macro-level hypotheses are tested in an experiment where individual, majority, and unanimity decision rules for punishing and rewarding are employed in an incentivized manner.

# 2 THEORY

## 2.1 The Public Goods Game

The Public Goods Game (PGG) is used as a model of public good problems. It is played by $n$ actors. All actors $i$ receive an endowment $w$. They simultaneously and independently decide whether to keep this endowment for themselves or contribute an amount $g_i \in [0, w]$ to a "group account". The total amount contributed to the group account by all n actors together, $g = \sum g_i$, is multiplied by a number $m$, with $1 < m < n$, and $mg$ is divided equally among all actors. Because $m < n$, the individual return obtained from the amount contributed to the group account is smaller than when it would have been kept to oneself ($mg_i/n < g_i$). Therefore, when the PGG is played once under standard game-theoretic assumptions - that is, when actors are rational in maximizing utility and selfish in that utility equals own payoff - contributing nothing is a dominant strategy, yielding the highest utility regardless what others do. This results in the unique Nash equilibrium of no contributions. However, since $m > 1$ the joint group outcome $nw - g + mg = nw + (m-1)g$ is maximized when everybody contributes the full endowment. Every player would then be better off compared to when all contribute nothing ($mnw/n = mw > w$). Thus, individually rational behaviour leads to a Pareto-suboptimal outcome, making the PGG a social dilemma (Kollock, 1998).

## 2.2 Behaviour in the PGG

The prediction of complete free-riding is typically refuted in experimental research employing the PGG. Instead, contributions averaging 50 percent of the endowment are consistently observed in one-shot PGGs (Kocher et al., 2008; Walker and Halloran, 2004). Also in repeated PGGs where group composition changes after each round, which resembles a series of one-shot interactions (Fehr and Gächter, 2002), subjects initially contribute 50 percent on average. However, in subsequent rounds contributions gradually decline to very low levels (Andreoni, 1988; Fehr and Gächter, 2002; see Ledyard, 1995 for an overview).

Research explaining this declining contribution pattern focuses on non-standard utility as an alternative behavioural assumption. It has been empirically established that actors in the PGG can be classified in two main preference types (Fehr and Gintis, 2007). Actors of the first type are rational and selfish free-riders who never contribute to the public good. Actors of the second type are conditional cooperators who contribute more, the more they expect others to contribute (Fischbacher et al., 2001; Ones and Putterman, 2007). These actors are assumed to derive utility from reciprocating others' expected contribution even in one-shot settings (Fehr and Gintis, 2007; Ones and Putterman, 2007; Ostrom, 2000). Conditional cooperators are heterogeneous in the extent to which they match others' contributions (Ostrom, 2000). Many are 'imperfect' reciprocators in that they contribute slightly below what they expect others to contribute on average (Fischbacher and Gächter, 2010). In an experiment specifically designed to identify preference types, Fischbacher et al. (2001) classify 50 percent of their subjects as conditional cooperators and 30 percent as free-riders.[1]

In repeated PGGs, conditional cooperators adapt their expectation of others' contribution on the basis of their experience of the average group contribution in the previous rounds (Fehr and Gintis, 2007; Fischbacher and Gächter, 2010). The more free-riders and imperfect conditional cooperators there are, the lower group contribution will be. Conditional cooperators decrease their contribution accordingly, which causes the average to further decline. This explains the decrease of cooperation over time (Fischbacher and Gächter, 2010).

### *2.3 The PGG with sanctions*

Sanctioning can be modelled by adding a second stage to the standard PGG. After all actors $i$ have determined their contribution and observed the contributions of the other group members, they decide for every other group member $j$ whether to pay an amount to punish and/or reward this actor. Let $s_{ij}$ denote the amount actor $i$ uses to sanction actor $j$. We assume here that an actor can only choose whether or not to sanction, but not the magnitude of the sanction: $s_{ij}$ is either a fixed amount $f > 0$ or zero. When the amount is used for punishment, a multiple $k$ of $f$ is subtracted from the payoff actor $j$ obtained in the PGG. The same amount is added to the payoff of actor $j$ when $s_{ij}$ is used for reward. Thus, in addition to the payoff from the standard PGG, every actor $j$ loses a total amount $k \sum_i s_{ij}$ of received punishment from all other actors $i$ or gains this amount of received rewards. Moreover, every actor $i$ forfeits $\sum_j s_{ij}$ by assigning sanctions to other actors $j$. This is the standard manner in which sanctions are executed in the PGG, denoted here as an IDR.[2]

In sanctioning systems with a CDR, all actors $i$ likewise decide whether to pay an amount to sanction others. Sanctioning under a CDR is different from an IDR in the sense that a sanction is only implemented when at least a proportion $p$ of actors sanctions the same recipient. Because we fixed the sanctioning amount $s_{ij}$ to 0 or $f$, this

[1] Virtually all remaining subjects were characterized as 'triangle' contributors (Fischbacher et al., 2001). These actors fully reciprocate others' expected contribution at 50 percent of the endowment, but their contribution declines when they expect others to contribute either more or less than this threshold.

[2] Note that details of this procedure can vary. For example, in many studies the amount $s_{ij}$ used to sanction can be chosen freely by actors between 0 and some numerical value.

implies that sanctions under a CDR are more severe than those under an IDR assigned by a smaller number of actors. Thus, every actor $j$ loses an amount $k \sum_i s_{ij}$ in a punishment system with a CDR when the proportion $q_j$ of actors $i$ for whom $s_{ij} = f$ is larger than or equal to $p$. The same applies to the amount gained under rewards. If $q_j < p$ no sanction is executed, that is, actor $j$ does not gain or lose money due to received sanctions. Moreover, the actors who proposed to sanction actor $j$ do not pay the cost of sanctioning if $q_j < p$. Thus, every actor $i$ who sanctions $j$ loses an amount $\sum_{j:q_j \geq p} s_{ij}$.

Actors are not informed about sanctions that were proposed by others but were not executed. Non-executed sanctions can therefore not influence behaviour of other actors than the ones who proposed the sanction. We assume one-shot interactions. Thus, actors cannot benefit from group members who increase their contribution in subsequent interactions after being sanctioned. This implies that long-term incentives for sanctioning, which are different under IDRs and CDRs, are ruled out. Moreover, non-executed sanctions are costless and not communicated to other group members. Therefore, actors have no incentive to take the probability that the sanction is executed into account when deciding whether or not to sanction under a CDR. Given these characteristics of the interaction situation, there is no reason to assume that actors make different sanctioning decisions under IDRs and CDRs.

We proceed with a review of empirical evidence and a theoretical account of contributing and sanctioning behaviour in the PGG with an IDR. This reveals which actors allocate sanctions, and which behaviours are more likely to be sanctioned. Multiple individual sanctions for a given behaviour imply a high consensus. Thus, given that the decision rule will not directly influence sanctioning decisions, those behaviours that are sanctioned individually by many are more likely to be sanctioned when a CDR is used. Behaviour in the PGG with an IDR then allows predicting the likelihood that sanctions will be implemented under a CDR.

## *2.4 Behaviour in the PGG with sanctions under an IDR*

As explained above, in one-shot PGGs and repeated PGGs with changing partners, actors derive no future benefits from current sanctioning. Therefore, rational selfish actors do not sanction when this is costly. Accordingly, the Nash equilibrium of the one-shot PGG with sanctions under standard assumptions of rationality and selfishness is no sanctioning and no contributions.

Despite this prediction, empirical evidence shows that actors frequently use punishment under an IDR in one-shot settings. It is consistently found that punishment is assigned in accordance with enforcing cooperation. That is, actors receive more punishment the less they contribute (Carpenter and Matthews, 2009; Casari and Luini, 2009), and the less they contribute compared to the average contribution of the group (Carpenter and Matthews, 2009; Fehr and Gächter, 2000; Ones and Putterman, 2007; Sefton et al., 2007). This punishment is mostly executed by high contributors (Decker et al., 2003; Sefton et al., 2007). It is also observed, however, that low contributors occasionally punish above-average contributors. This 'antisocial' punishment is usually carried out by a small number of actors (Casari and Luini, 2009). The extent to which it occurs varies greatly between subject pools (Herrmann et al., 2008). The effect of cooperation-enforcing and antisocial punishment differs. Below-average contributors increase their contribution in the subsequent round after being punished (Fehr and Gächter, 2002), but for above-average contributors empirical evidence is mixed. Some studies show that above-average contributors decrease their contribution after being sanctioned (Bochet et al., 2006; Masclet et al., 2003; Ones and Putterman, 2007); others find no effect of antisocial punishment on contribution (Denant-Boemont et al., 2007).

Like punishments, rewards are typically used to enforce cooperation in one-shot settings. High contributors tend to reward other high contributors (Andreoni et al., 2003; Walker and Halloran, 2004). However, although rewards are allocated to above-average contributors, it is not so clear as it is for punishment that the amount of rewards received increases with the (positive) deviation from the average group contribution (Sefton et al., 2007; Walker and Halloran, 2004). In repeated PGGs with fixed group composition it is found that rewards are frequently used in every successive interaction (Rand et al., 2009; Milinski and Rockenbach, 2011). However, the use of rewards declines over time in fixed groups when actors cannot infer who rewarded them (Sefton et al., 2007).

## *2.5 Non-selfish utility in the PGG with sanctions*

Rational selfish free-riders never sanction when this is costly. However, anticipation on being sanctioned will induce them to contribute, provided that the loss due to received punishment or gain from rewards offsets the payoff advantage of free-riding (Fehr and Fischbacher, 2004).

Non-selfish actors could derive utility from sanctioning even in one-shot interactions (Diekmann and Voss, 2003). Empirical evidence shows that sanctions are mostly allocated by high contributors. However, it remains an open question to what extent these actors are also conditional cooperators in the standard PGG (Ones and Putterman, 2007). According to the theory of strong reciprocity, conditional cooperators reciprocate contributions of others not only through matching their expected contribution, but also through sanctioning (Fehr and Gintis, 2007). Others regard actors who derive utility from sanctioning as a separate behavioural type, which might have overlap with conditional cooperators (Ostrom, 2000).

Empirical evidence is indeed consistent with the assumption that people derive utility from punishing and rewarding in one-shot settings. Fehr and Gächter (2002) already noted that subjects experience anger when they observe free-riding in a hypothetical situation. This anger increases the more the free-rider deviates from the average of others. In a neurobiological experiment, De Quervain et al. (2004) show that the human reward system is activated in the brain of an actor punishing a defector. Utility from rewarding is addressed by Dawes et al. (2007), who conducted an experiment in which subjects could decide on a costly in- or decrease of a random amount of tokens other subjects had received. They found that subjects who afterwards indicated more anger and annoyance towards those with a high amount also spent more to increase low and reduce high amounts received by others. Yet, despite utility derived from sanctioning, it is found that actors sanction less the higher the costs of sanctioning are (Carpenter, 2007; Fehr and Fischbacher, 2004). Thus, actors take their own payoff into account in sanctioning decisions.

As mentioned above, some actors use sanctions antisocially. Although they are relatively rare, antisocial sanctioners constitute a separate type of actors. These actors free-ride in the PGG, and subsequently punish high contributors (Herrmann et al., 2008). A motive for antisocial punishment might be revenge on previous punishment received from high contributors, a desire to increase relative payoff advantage of free-riding, or a dislike of do-gooders (Gächter and Herrmann, 2009). Alternatively, it could be that actors occasionally punish high contributors by mistake. Antisocial rewards, i.e. rewarding of low contributors, has to our knowledge never been explored in previous research. However, antisocial rewards increase the payoff discrepancy between high and low contributors, and can thus have the same detrimental effect on cooperation as antisocial punishment.

Punishment and reward are used in different ways. The possibility of being punished might be enough to deter free-riding, such that there is no need to allocate punishment. However, every time an actor makes a high contribution, rewards actually have to be carried out to induce free-riders to contribute (Dari-Mattiacci and De Geest, 2010). Thus, when contributions in a population increase due to the existence of a sanctioning system, more rewards than punishments have to be allocated. In one-shot settings, actors cannot establish a norm of direct mutual rewarding. They are therefore unsure whether the costs of allocating rewards will be offset by reciprocation (Rand et al., 2009). This makes rewarding more expensive than punishment in the one-shot PGG. As stated above, more expensive sanctioning implies that fewer sanctions are assigned. This explains why, without opportunities for directly reciprocating received rewards, actors initially attempt to reward but eventually give up when others do not continue to reward as well.

## *2.6 Micro-level hypotheses*

Before turning to differences in contribution levels between IDRs and CDRs, we capture the theory developed on contributing and sanctioning behaviour in a number of hypotheses. These hypotheses will be used as a micro-level framework summarizing which actors are likely to sanction, and how actors react to receiving cooperation-enforcing or antisocial sanctions. When theorizing about the effect of sanctioning decision rules on contributions, we assume that actors behave as summarized in this framework.

We first derive hypotheses sanctioning behaviour. Although antisocial punishment is sometimes observed, punishment is usually allocated by cooperation-enforcing high contributors. Accordingly, we hypothesize that actors are more likely to punish others the more they contributed themselves.

**Hypothesis 1:** The more an actor contributes, the higher this actor's likelihood to assign punishment.

Punishment of high contributors is more often targeted at free-riders than that of low contributors, who might punish antisocially. Thus, the more an actor contributed the more likely he is to punish a free-rider. This implies that we expect an interaction between the contribution of the actor allocating punishment and the contribution of the recipient on the likelihood to sanction. We argue that actors perceive free-riding both in the sense of the recipient contributing a low amount and in the sense of contributing less than the other group members. This means that low as well as below-average contributors are likely to be punished by high contributors.

**Hypothesis 2a:** The more an actor contributes, the more this actor's likelihood of assigning punishment decreases with the contribution of the sanction's recipient.

**Hypothesis 2b:** The more an actor contributes, the more this actor's likelihood of assigning punishment increases with the negative deviation of the sanction's recipient from the group average contribution.

Also reward is predominantly allocated by high contributors.

**Hypothesis 3:** The more an actor contributes, the higher this actor's likelihood to assign reward.

High contributors are more likely to reward other high contributors. This applies both in absolute sense, and compared to the average of other group members. Again, we hypothesize an interaction between the contribution of the rewarding actor and the contribution of the recipient.

**Hypothesis 4a:** The more an actor contributes, the more this actor's likelihood of assigning reward increases with the contribution of the sanction's recipient.

**Hypothesis 4b:** The more an actor contributes, the more this actor's likelihood of assigning reward increases with the positive deviation of the sanction's recipient from the group average contribution.

Unlike punishments, in order to enforce cooperation rewards have to be allocated every time an actor makes a high contribution. They are therefore costly to maintain when direct reciprocation is impossible. Accordingly, the likelihood of rewarding decreases over rounds.

**Hypothesis 5:** The more rounds have already been played, the lower the likelihood that rewards are allocated.

We now turn to the effect of sanctions on contribution. Receiving punishment leads to conformation to behaviour of other actors, in order to avoid receiving punishment in future interactions. Free-riders thus increase and high contributors decrease contribution after being punished. Consequently, their contribution is more in line with others' average.

**Hypothesis 6:** The more an actor contributing below the average is punished, the more this actor contributes in the subsequent interaction.

**Hypothesis 7:** The more an actor contributing above the average is punished, the less this actor contributes in the subsequent interaction.

Rewards reinforce current behaviour. Above-average contributors will thus contribute more and below-average contributors less when they are rewarded, provided they did not already contribute the full endowment or free-ride completely, respectively.

**Hypothesis 8:** The more an actor contributing above the average is rewarded, the more this actor contributes in the subsequent interaction.

**Hypothesis 9:** The more an actor contributing below the average is rewarded, the less this actor contributes in the subsequent interaction.

## *2.7 Macro-level effects of CDRs*

Only the sanctions on which required consensus is reached are executed under a CDR. Given sanctioning behaviour as predicted in the micro-level hypotheses, it is likely that there will be more consensus on some sanctions than on others. This gives rise to different contribution levels under IDRs versus CDRs. Macro-level hypotheses differ for punishment and reward.

Under an IDR, all allocated punishments are carried out. This implies that high contributors will frequently punish free-riders. Free-riders will receive more punishment the less they contribute in absolute sense and compared to the others. Also, antisocial punishers have the opportunity to punish high contributors.

The situation is different when only those sanctions are implemented to which a majority of actors consents. A large proportion of actors derives utility from sanctioning. It is therefore likely that often majority consent is reached on punishment of free-riders. The more a free-rider deviates from the average, the higher the chance that consent will be reached. Conversely, antisocial punishment is relatively rare. Therefore, it will be unlikely that a majority of actors agrees on punishing a high contributor, avoiding the negative effects of antisocial punishment. Thus, a majority sanctioning system will rule out antisocial punishment while at the same time cooperation-enforcing punishment is likely to be implemented. We therefore expect a majority decision rule to lead to higher contribution levels than an IDR.

**Hypothesis 10a:** Contribution is higher under a majority than under an individual punishment decision rule.

Some previous studies indeed found that majority consent is sufficient to rule out antisocial punishment, but that cooperation-enforcing punishment could still be implemented. Casari and Luini (2009) found that punishment was more effective when two out of four actors had to agree on sanctioning a fifth. Antisocial punishment was to a large extent ruled out under this decision rule. Likewise, Ertan et al. (2009) let subjects choose whether or not to enable punishment of high contributors. While this was sometimes favoured by a number of free-riders, it was never implemented because a majority opposed the possibility.

Under a unanimity decision rule punishment is only executed when all remaining group members decide to punish an actor. Antisocial punishment is therefore even less likely than under a majority decision rule. However, also for cooperation-enforcing punishment a unanimity decision rule requires a very high proportion of actors willing to punish. Therefore, it will be difficult to implement any punishment at all. Conversely, under an IDR there could be antisocial punishment, although the vast majority of punishment should be targeted at below-average contributors. It is therefore likely that contribution levels under a unanimity punishment decision rule are lower than under an individual rule.

**Hypothesis 10b:** Contribution is higher under an individual than under a unanimity punishment decision rule.

As explained above, continuous need of rewarding makes reciprocating through rewards more expensive than through punishment, which causes the use of rewards to decline (Dari-Mattiacci and De Geest, 2010). Thus, more punishment than reward will be executed under every decision rule, making sanctioning through punishment more effective.

Therefore, we argue that for every decision rule contribution is higher under punishment than under reward.

**Hypothesis 11:** For every decision rule, contribution is higher under punishment than under reward.

The more actors are required for a reward to be executed, the more likely it is that too many actors give up on using rewards. Thus, the more actors are required the more likely it is that consensus cannot be reached anymore. Also, antisocial rewards have to be carried out every time an actor free-rides in anticipation on being rewarded. Antisocial rewards are thus likewise costly to maintain. Therefore, while antisocial rewards might be occasionally allocated it is unlikely that they are persistently problematic for enforcing cooperation. Thus, rewarding under an IDR is not thwarted by antisocial sanctions as much as punishment, while it is difficult to raise enough actors to agree on rewards under a CDR. The more actors are required to agree, the more problematic enforcing cooperation becomes. Accordingly, we hypothesize that the more actors are required to agree on rewarding, the less rewards will be carried out and the lower contribution levels are.

**Hypothesis 12a:** Contribution is higher under an individual than under a majority rewarding decision rule.

**Hypothesis 12b:** Contribution is higher under a majority than under a unanimity rewarding decision rule.

# 3 EXPERIMENTAL DESIGN

In the experiment, subjects participated in interaction situations based on the PGG as described above with group size $n = 4$; endowment $w = 20$, and multiplier $m = 1.6$. The outcome of the game represented points that subjects earned. After the experiment, subjects received one eurocent for every sixty points earned.

The experiment comprised three parts. In the first part, preferences for conditional cooperation were assessed using a measure designed by Fischbacher et al. (2001). First, subjects decided on an unconditional contribution, i.e. how much to contribute in the PGG in a group with three other subjects. Second, subjects made this same decision conditional on others' average contribution. Thus, they decided how much they would contribute for every possible average of the three other group members (strategy method, Selten, 1967). The more conditionally cooperative a subject is, the more contribution should increase with others' average. Subjects were randomly matched in groups of four. For three randomly chosen group members, payoff was calculated based on the unconditional contribution. For the fourth group member the conditional contribution corresponding to the average unconditional contribution of the three others was used. This makes both decisions incentive-compatible. Note that conditionally cooperative preferences were always assessed at the beginning of a session, prior to playing the actual PGGs. Fischbacher and Gächter (2010) measured conditional cooperation using a similar design, administered either at the start or end of the experiment. They did not find a sequence effect, suggesting that measuring preferences does not significantly influence subsequent behaviour.

In the second part of the experiment, the standard PGG as described above was played for ten rounds. Between the rounds, subjects were randomly rematched into different groups. They could not infer their group members' previous decisions. After every round, subjects were informed about the contribution of the others in their group and their own payoff.

In the third part, the PGG with sanctions was employed. In every session, ten rounds were played with only punishment and ten rounds with only reward; the order varied between sessions. Both reward and punishment took place in one of three experimental conditions; individual, majority, or unanimity. In all three conditions, subjects first decided upon a contribution. Subsequently, they were informed about contributions of their group members and decided for all three others separately whether to sanction this person. If executed, a sanction added or subtracted six points from the earnings of the recipient at a cost of two points. This cost ratio of 1:3 is often used in PGG experiments (cf. Fehr and Gächter, 2002). The effect and cost of the sanction were chosen to ensure that receiving a sanction has a severe impact on payoffs. Because the amount by which actors could sanction was fixed, the severity of the sanction is equal to the number of actors sanctioning.

In the individual condition, all assigned rewards and punishments were implemented. Subjects who received multiple sanctions were sanctioned by the cumulative amount while all subjects allocating the sanction paid the cost of two points. The procedure in the majority condition was exactly the same, except that the sanction was only executed when at least two group members wanted to sanction the same recipient. Thus, an actor sanctioned by two others then lost twelve points, while both sanctioning actors lost two points. In the unanimity condition, the sanction was only executed when it was requested by all three remaining group members. When the number of subjects who wanted to sanction was insufficient in the majority or unanimity condition, the sanction was not executed and no costs had to be paid. After each round, subjects were informed about all sanctions that had been executed in their group but could not infer who allocated them. No information was provided about sanctions that were not executed. Again, subjects were randomly rematched between the rounds.

The experiment was programmed using z-Tree (Fischbacher, 2007) and conducted at the laboratory of Utrecht University. Subjects were recruited using the online recruiting system ORSEE (Greiner, 2004). Twelve experimental sessions were held, four in each experimental condition of which two with reward first and two with punishment first. Instructions were provided on paper. It was made clear that the instructions were always truthful and identical for all subjects in a session. In the first set of instructions, the standard PGG and the first two parts of the experiment were explained. It was announced that there would be further tasks, but not what these tasks entailed. These instructions included a number of control questions, which appeared on the computer screen. When a subject did not answer correctly to a question, the answer was explained on the screen. Additional instructions, adapted for each experimental condition, were provided for the reward as well as for the punishment part. The options in the PGG were labelled in a neutral way: punishment and reward were called 'subtracting' and 'adding' points, respectively.

A total number of 184 student subjects participated in the experiment (32 percent male; 34 percent economics major). Payoffs averaged €12.50, with a minimum of €8.50 and a maximum of €15.

# 4 METHOD AND RESULTS

## *4.1 Descriptive results*

Figure 1 shows the average contributions in the PGGs over the rounds in the baseline and in each experimental condition. Note that all subjects participated first in the baseline, and subsequently in reward as well as punishment of one of the conditions.

Contributions are initially around fifty percent of the endowment. This is in line with consistent evidence from previous research (Ledyard, 1995). After the first round, Figure 1 shows strong differences in contribution levels between the conditions. Contributions in the baseline decline to almost zero. Conversely, individual and majority punishment are the only conditions under which contributions increase over time. Under all decision rules punishment results in higher contributions than reward. For both reward and punishment the individual and majority conditions lead to higher contributions than unanimity.

When a subject was punished in the majority condition, in fifty-seven percent of the cases this was by one person only and therefore the punishment was not carried out. Likewise, in eighty-one percent of the cases in which a subject was sanctioned in the unanimity condition the required number of three sanctioning subjects was not reached.

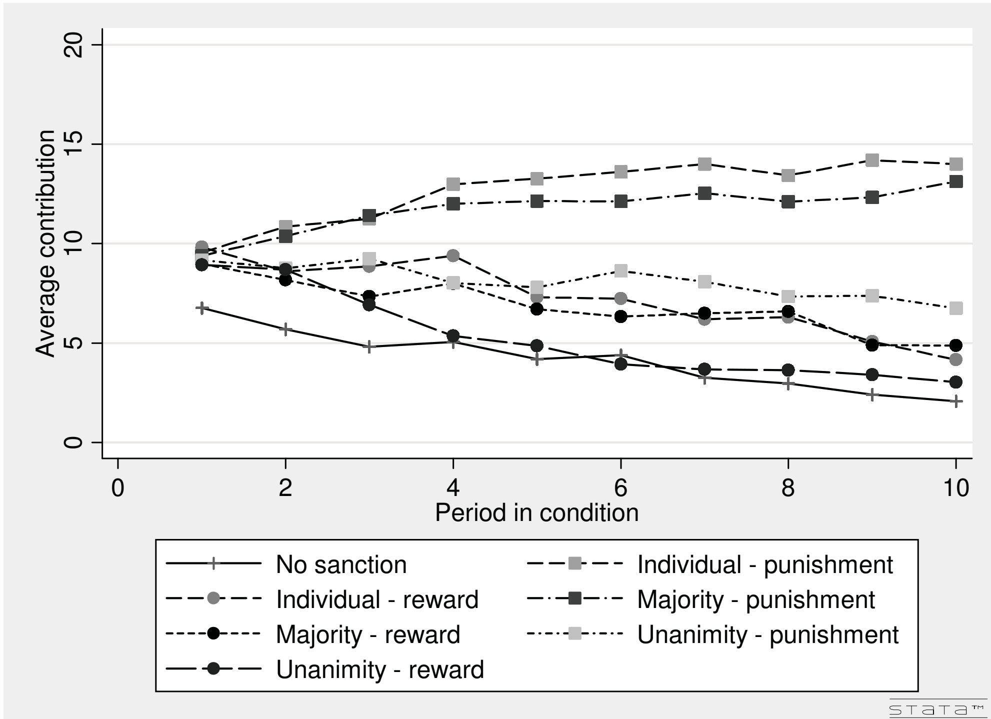


Figure 1: *Average contribution in the PGGs, separated for each round and experimental condition*

For reward, seventy-two percent of sanctions in the majority condition and ninety-seven percent of sanctions in the unanimity condition were not implemented.

Figure 2 shows the average number of sanctions allocated and average number of sanctions carried out for different deviations of the recipient from the average contribution of the other group members. Note that between one and three other group members can propose to sanction. Figure 2 shows a clear trend of more punishment proposed on average the more the recipient negatively deviates from the average contribution of others. There is also some punishment visible of above-average contributors, primarily in the majority and unanimity conditions. However, especially the antisocial sanctions seem to be ruled out by the punishment systems with a CDR. Furthermore, Figure 2 shows that more rewarding is proposed for high contributors, but not so clearly that more rewards are proposed the higher the contribution.

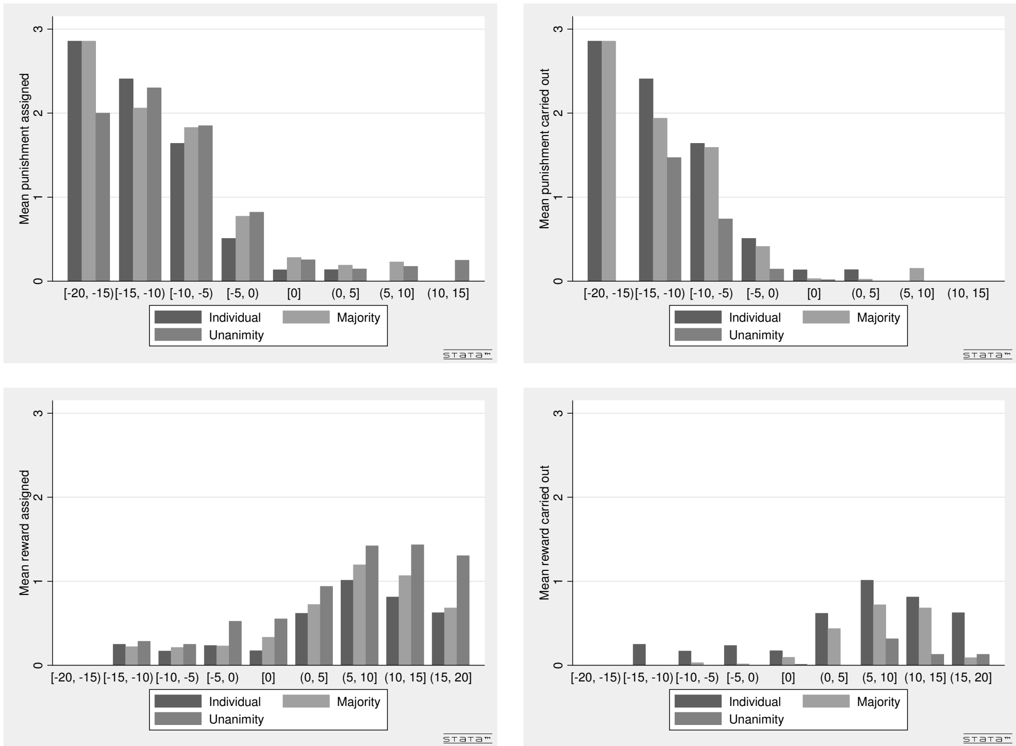


Figure 2: *Average punishment (above) and reward (below) assigned (left) and carried out (right) for different deviations from the average contribution of other group members, separated for each experimental condition*

## *4.2 Contribution - methods*

The first dependent variable, contribution, is measured as the contribution decisions of subjects in the PGG. First, we test macro-level hypotheses by comparing dummies for the experimental conditions individual, majority, and unanimity punishment and reward. Second, we test the micro-level hypotheses explaining differences between experimental conditions. Punishment and reward conditions are analyzed separately.

In the micro-level models, sanctions received is measured as the number of others who had sanctioned the subject in the previous round. Only executed sanctions are included. Furthermore, three dichotomous variables indicate whether in the previous round a subject had contributed more than five points below the average of other group members, more than five points above the average, or did not deviate from the average by more than five points. These three dummies for previous deviation are interacted with the number of sanctions received to test whether the effect of being sanctioned is different for above- and below-average contributors.

Previous deviation was measured using dummies rather than a continuous variable indicating the precise extent of the deviation. This is because a continuous variable interacted with received reward would test if subjects increase (decrease) their contribution more, the higher (lower) the contribution for which they were rewarded. This is unrealistic, since contribution is limited between zero and twenty. The boundaries of five points from the average are chosen such that the deviation is substantial enough for subjects to perceive sanctions as clearly norm-enforcing or antisocial. Accordingly, log likelihood of models in which these boundaries were varied was equal to or lower than that of the models presented here. This suggests that subjects indeed react differently to receiving sanctions depending on whether or not they deviate more than five points.

We control for the subjects' contribution in the previous round, round number, and experimental condition. Furthermore, preference for conditional cooperation is included, measured as the slope of the conditional contribution assessed in the first part of the experiment. The steeper the slope, the more a subject indicated to contribute more when others do so as well.[3] All control variables are centred.

We use Tobit regression to take into account that contribution has a limited range, between zero and twenty, of which both extremes are often chosen. The unit of analysis are decisions in the PGGs. Random effects at the subject level are included to model that every subject makes multiple contribution decisions. Apart from this nesting of decisions in subjects, within a session subjects often encounter others with whom they or their group members have interacted previously. Thus, subjects are interdependent within sessions. It is not possible to include both the subject and session level in a Tobit model. Therefore, all models were replicated using multilevel regression, in which the subject and session level are included but where contribution is treated as if its range is unlimited. Also, we estimated the models using Tobit regression with random effects at the session level to test if disregarding this level in the models presented below influenced the results. Results are robust in these alternative analyses unless stated otherwise.

## *4.3 Contribution – results*

Table 1 shows differences in contribution decisions between the experimental conditions. The baseline condition, in which every subject participated, serves as a reference. Contributions in all experimental conditions except unanimity reward were higher than in the baseline. Contrary to Hypothesis 10a, contribution under punishment is higher in

[3]Two subjects whose slopes are zero, but who do make positive conditional contributions are excluded from the analysis. A zero slope thus indicates a preference for unconditional free-riding.

the individual than the majority condition ($\chi^2(1) = 29.51$; $p < 0.01$). The other macro-level hypotheses are confirmed. Contribution under punishment is higher in the individual than the unanimity condition ($\chi^2(1) = 136.58$; $p < 0.01$), confirming Hypothesis 10b. As predicted in Hypothesis 11, contribution is higher under punishment than reward in the individual ($\chi^2(1) = 228.83$; $p < 0.01$), majority ($\chi^2(1) = 246.01$; $p < 0.01$) and unanimity ($\chi^2(1) = 122.01$; $p < 0.01$) condition. Finally, contribution under reward is higher in the individual than the majority condition ($\chi^2(1) = 23.76$; $p < 0.01$) and higher in the majority than the unanimity condition ($\chi^2(1) = 12.79$; $p < 0.01$). This confirms Hypotheses 12a and 12b.

Table 1: *Tobit regression on contribution decisions with random effects at subject level (5460 decisions, of which 2376 censored, by 182 subjects)*

| | Model 1 | |
|---|---|---|
| | Coeff. | Sd. |
| Punishment - individual | 13.938** | 0.518 |
| Punishment - majority | 10.239** | 0.464 |
| Punishment - unanimity | 5.866** | 0.479 |
| Reward - individual | 6.184** | 0.528 |
| Reward - majority | 2.770** | 0.474 |
| Reward - unanimity | 0.340 | 0.501 |
| Constant | 0.786 | 0.522 |
| $\sigma_u$ | 6.372** | 0.380 |
| $\sigma_e$ | 7.934** | 0.113 |
| Log Likelihood | -12773.784 | |

*Significant at .05-level; **Significant at .01-level (2-sided)

The micro-level model for the punishment conditions is presented in Table 2. Only main effects are included in Model 2. Several control variables are significant. Contribution is lower in the unanimity compared to the individual condition, and higher the more a subject contributed in the previous round. The difference between the individual and majority condition is not significant in this model. Subjects who contributed 5 points or more below the average increase their contribution compared to around-average contributors. However, this effect was no longer significant when the regression was modelled as multilevel or with random effects at session level. Therefore, we cannot interpret the effect. Note that no hypothesis was formulated on this coefficient. Subjects who contributed above the average decrease their contribution compared to around-average contributors. Also, contribution is higher the more punishment was received previously.

Interaction effects are included in Model 3. The main effect of punishment is excluded from this model, so that the three interactions represent the effect of received punishment for the three groups of subjects belonging to specific deviations from the mean contribution. The model shows that subjects contributing below the average increase their contribution the more they are punished. Hypothesis 6 is thus confirmed. The insignificant main effect of negative deviation indicates that subjects who contributed below the average but were not punished do not significantly increase their contribution compared to around-average contributors. Subjects who contributed above the average decreased their contribution if they had not been punished, but did not decrease their contribution further after receiving punishment. Thus, no support is found for Hypothesis 7.

Table 2: *Tobit regression on contribution decisions in the punishment conditions with random effects at subject level (1638 decisions, of which 345 censored, by 182 subjects)*

| | Exp. | Hyp. | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|
| | dir. | nr. | Coeff. | Sd. | Coeff. | Sd. |
| Previous punishment received | | | 1.095** | 0.166 | | |
| Prev. neg. deviation > 5 | | | 0.966* | 0.477 | -0.243 | 0.593 |
| × Punishment received | + | 6 | | | 1.750** | 0.252 |
| Prev. dev. ≤ 5 | | | ref. | | ref. | |
| × Punishment received | | | | | 0.650** | 0.215 |
| Prev. pos. deviation > 5 | | | -1.672** | 0.382 | -1.714** | 0.387 |
| × Punishment received | - | 7 | | | -0.015 | 1.159 |
| Previous contribution | | | 0.635** | 0.046 | 0.644** | 0.046 |
| Slope conditional contribution | | | 0.963 | 0.564 | 0.893 | 0.546 |
| Period | | | -0.016 | 0.043 | -0.022 | 0.043 |
| Individual | | | ref. | | ref. | |
| Majority | | | -0.183 | 0.618 | -0.205 | 0.598 |
| Unanimity | | | -3.047** | 0.659 | -2.983** | 0.642 |
| Constant | | | 9.916** | 0.524 | 9.534** | 0.525 |
| $\sigma_u$ | | | 2.986** | 0.272 | 2.866** | 0.273 |
| $\sigma_e$ | | | 4.225** | 0.094 | 4.223** | 0.094 |
| Log Likelihood | | | -4141.843 | | -4135.603** | |

*Significant at .05-level; **Significant at .01-level (2-sided)

Model 4 in Table 3 shows the determinants of contribution decisions in the reward conditions. In this model the differences between experimental conditions are not significant.

The other control variables are significant; contribution is higher the more conditionally cooperative a subject is, and decreases over rounds. Again, the effect of previous negative deviation is no longer significant in multilevel regression or with random effects at session level. Subjects who previously contributed above the average decrease their contribution compared to around-average contributors. Finally, the more rewards a subject had previously received, the higher the contribution.

In Model 5, the interaction effects are included. Again, the three interactions represent the separate main effects. This shows that subjects who had contributed above the average significantly increase their contribution after receiving reward. Only if they were not rewarded, contribution decreased. This confirms Hypothesis 8. Very few subjects received rewards after a below-average contribution. Hence, we find no significant effect of being rewarded for around-average or below-average contributors. Hypothesis 9 is not confirmed.

Table 3: *Tobit regression on contribution decisions in the reward conditions with random effects at subject level (1638 decisions, of which 345 censored, by 182 subjects)*

| | Exp. | Hyp. | Model 4 | | Model 5 | |
|---|---|---|---|---|---|---|
| | dir. | nr. | Coeff. | Sd. | Coeff. | Sd. |
| Previous reward received | | | 1.591** | 0.513 | | |
| Prev. neg. deviation > 5 | | | 1.800* | 0.889 | 1.627 | 0.903 |
| × Reward received | - | 9 | | | 1.738 | 2.032 |
| Prev. dev. ≤ 5 | | | ref. | | ref. | |
| × Reward received | | | | | -0.193 | 0.874 |
| Prev. pos. deviation > 5 | | | -4.837** | 1.001 | -6.003** | 1.103 |
| × Reward received | + | 8 | | | 2.355** | 0.613 |
| Previous contribution | | | 0.827** | 0.089 | 0.854** | 0.090 |
| Slope conditional contribution | | | 5.528** | 1.597 | 5.511** | 1.612 |
| Period | | | -0.770** | 0.124 | -0.777** | 0.124 |
| Individual | | | ref. | | ref. | |
| Majority | | | 0.298 | 1.736 | 0.249 | 1.755 |
| Unanimity | | | -3.008 | 1.780 | -3.139 | 1.799 |
| Constant | | | 3.535** | 1.335 | 3.578** | 1.374 |
| $\sigma_u$ | | | 8.349** | 0.787 | 8.453** | 0.790 |
| $\sigma_e$ | | | 9.630** | 0.318 | 9.576** | 0.316 |
| Log Likelihood | | | -3069.150 | | -3066.027* | |

*Significant at .05-level; **Significant at .01-level (2-sided)

## 4.4 *Sanctioning - methods*

The second dependent variable in the analysis of the micro-level framework are the decisions whether or not to sanction. These are three observations for each subject in each period; one for every other group member.

The first independent variable is a subjects' own contribution. Furthermore, contribution of the recipient is included as a continuous variable. Deviation of the recipient from the average of others is measured as the contribution of the recipient minus the average of the other group members. The variable positive deviation includes all positive values of this measure, negative values are set to zero. Absolute negative deviation represents the extent of the deviation of all negative values, zero for positive deviations. For punishment, the contribution and absolute negative deviation of the recipient are interacted with the subjects' own contribution to test whether high contributors are more likely to punish the less the recipient contributes, and the further he deviates from the average. For reward, contribution and positive deviation of the recipient are interacted with subjects' contribution. We control for experimental condition, slope of the conditional contribution, and for sanctions assigned and received by the subject in the previous round.

We use logistic regression to analyze the dichotomous sanctioning decisions. Every subject makes three sanctioning decisions, one for every other group member, in all ten periods. Decisions are thus nested within periods, and within subjects. A multilevel intercept-only model with decisions nested in periods and subjects revealed that variance at the period level is negligible for both punishment and rewarding decisions. We therefore use multilevel models with decisions nested only in subjects.

## 4.5 *Sanctioning - results*

Models on punishment decisions are displayed in Table 4. Model 6 shows that there are no differences between the experimental conditions in the likelihood that a subject decides to punish another. We do find that subjects who have received or allocated punishment in the previous round are more likely to punish. The likelihood of punishing increases with contribution, confirming Hypothesis 1. Also, the more a recipient negatively deviates from others' contribution, the higher the likelihood that punishment is allocated while no effect is found for positive deviation. Finally, the more a group member contributes, the less likely a subject is to punish this person.

Model 7 shows a significant interaction effect of contribution with the contribution of the recipient, confirming hypothesis 2a. A significant interaction with negative deviation of

Table 4: *Multilevel logistic regression on decision whether to punish nested in subjects (4914 decisions by 182 subjects)*

| | Exp. | Hyp. | Model 6 | | Model 7 | |
|---|---|---|---|---|---|---|
| | dir. | nr. | Coeff. | Sd. | Coeff. | Sd. |
| Contribution | + | 1 | 0.091** | 0.018 | 0.121** | 0.019 |
| Contribution recipient | | | -0.213** | 0.027 | -0.160** | 0.028 |
| × Contribution | + | 2a | | | -0.020** | 0.003 |
| Positive dev. recipient | | | -0.021 | 0.032 | -0.078* | 0.034 |
| Absolute neg. dev. recipient | | | 0.290** | 0.029 | 0.217** | 0.033 |
| × Contribution | + | 2b | | | 0.013** | 0.004 |
| Round | | | 0.029 | 0.020 | 0.049** | 0.021 |
| Individual | | | ref. | | ref. | |
| Majority | | | 0.156 | 0.397 | 0.159 | 0.451 |
| Unanimity | | | -0.226 | 0.414 | -0.204 | 0.466 |
| Slope conditional contribution | | | 0.300 | 0.360 | 0.159 | 0.410 |
| Previous punishment received | | | 0.301** | 0.067 | 0.269** | 0.067 |
| Previous punishment assigned | | | 0.289** | 0.067 | 0.252** | 0.068 |
| Constant | | | -1.843** | 0.326 | -1.789** | 0.363 |
| | | | | | | |
| Subject level | | | | | | |
| Constant | | | 1.949** | 0.170 | 2.234** | 0.194 |
| | | | | | | |
| Log Likelihood | | | -1507.103 | | -1447.098** | |

*Significant at .05-level; **Significant at .01-level (2-sided)

the recipient confirms Hypothesis 2b. High contributors are thus more likely to punish the less a recipient contributes in absolute sense, and relative to the average of others.

Table 5 shows the models on rewarding. Main effects included in Model 8 show that subjects in the unanimity condition are more likely than in the individual condition to allocate rewards. Furthermore, subjects are more likely to reward the more rewards they had allocated in the previous period. The effect of period is significant, confirming Hypothesis 5. Also, subjects are more likely to reward the more the recipient contributes, but not the higher the positive deviation from the average. We do find that rewarding is less likely the more the recipient negatively deviates. Hypothesis 3 is supported: subjects who made a higher contribution are more likely to reward.

Model 9 shows the interaction of a subjects' own contribution with the contribution and positive deviation of the recipient. The significant effects indicate that high contribu-

Table 5: *Multilevel logistic regression on decision whether to reward nested in subjects (4914 decisions by 182 subjects)*

| | Exp. | Hyp. | Model 8 | | Model 9 | |
|---|---|---|---|---|---|---|
| | dir. | nr. | Coeff. | Sd. | Coeff. | Sd. |
| Contribution | + | 3 | 0.070** | 0.012 | 0.037** | 0.013 |
| Contribution recipient | | | 0.158** | 0.019 | 0.151** | 0.020 |
| × Contribution | + | 4a | | | 0.005** | 0.002 |
| Positive dev. recipient | | | 0.032 | 0.020 | 0.073** | 0.023 |
| × Contribution | + | 4b | | | 0.007** | 0.003 |
| Absolute neg. dev. recipient | | | -0.115** | 0.025 | -0.047 | 0.026 |
| Round | - | 5 | -0.052* | 0.023 | -0.064** | 0.024 |
| Individual | | | ref. | | ref. | |
| Majority | | | 0.338 | 0.500 | 0.350 | 0.487 |
| Unanimity | | | 1.354** | 0.504 | 1.300** | 0.491 |
| Slope conditional contribution | | | 0.747 | 0.451 | 0.664 | 0.438 |
| Previous reward received | | | -0.179 | 0.092 | -0.187 | 0.100 |
| Previous reward assigned | | | 0.347** | 0.075 | 0.351** | 0.076 |
| Constant | | | -3.504** | 0.384 | -3.449** | 0.374 |
| | | | | | | |
| Subject level | | | | | | |
| Constant | | | 2.437** | 0.210 | 2.354** | 0.205 |
| | | | | | | |
| Log Likelihood | | | -1331.108 | | -1281.803** | |

*Significant at .05-level; **Significant at .01-level (2-sided)

tors are more likely to reward the higher and the further above the average someone contributes, confirming Hypotheses 4a and 4b.

# 5 CONCLUSION AND DISCUSSION

In this paper we compared the effect of individual, majority, and unanimity decision rules for implementing punishment and reward on actors' ability to enforce cooperation in a Public Goods Game (PGG). For punishment, we conjectured that contributions are higher under a majority than an individual decision rule (Hypothesis 10a). However, we find higher contributions under the individual decision rule instead. As expected, we do find that contribution is lower under a unanimity than an individual punishment decision rule (Hypothesis 10b). For reward, the hypotheses concerning the effect of decision rules on contribution are confirmed. We find that contribution is higher under an individual than a majority decision rule (Hypothesis12a) and higher under a majority than a unanimity decision rule (Hypothesis 12b). In sum, for both punishment and reward contributions are lower, the more actors are required to agree on sanctioning. Also, for every decision rule contribution is higher under punishment than reward (Hypothesis 11).

Findings on individual behaviour, as captured in micro-level hypotheses, offer an explanation for the observed differences in contribution between decision rules. The emerging pattern is very similar for reward and punishment. Hypotheses on the use of cooperation-enforcing sanctions are all confirmed. High contributors are more likely to punish (Hypothesis 1) and to reward (Hypothesis 3) than low contributors. These high contributors enforce the norm that others should contribute as well. That is, they are more likely to punish the less a recipient contributes (Hypotheses 2a) and the lower the contribution of the recipient is compared to the other group members (Hypothesis 2b). Likewise, high contributors reward group members who also make a high contribution (Hypothesis 4a) and who contribute more compared to the others (Hypothesis 4b). In other words, there is more consensus on sanctions among high contributors, the more an actor violates or adheres to their cooperative norm. Still, many punishments and rewards under the majority and unanimity decision rules were not executed. This implies that reaching the required number of actors was difficult despite the high consensus on whom to target.

When low contributors are punished, they contribute more in the subsequent interaction (Hypothesis 6). Similarly, actors who are rewarded for contributing more than other group members increase their contribution (Hypothesis 8). Thus, as hypothesized we find strong evidence that cooperation-enforcing sanctions have a positive effect on contributions. Conversely, antisocial sanctioning occurred too infrequently to affect contribution levels. We cannot confirm that high contributors decrease their contribution after being punished antisocially (Hypothesis 7). Likewise, contrary to our expectations, free-riders who are rewarded antisocially do not decrease their contribution further (Hypothesis 9).

In sum, we find strong evidence for cooperation-enforcing sanctions, and their positive effects on contribution. Concurrently, antisocial sanctions occur too infrequently to affect cooperation. This makes an IDR unproblematic: punishment is targeted at cooperators regardless of the possibility for individual actors to sanction antisocially. Because more cooperation-enforcing sanctions are obstructed the more actors are required for the collective decision rule (CDRs), we observe lower contribution levels the more actors are required to agree. The observed micro-level behaviour thus explains the macro-level finding of lower contribution levels under unanimity than majority, and lower contributions in the majority than in the individual condition.

The use of rewards decreases over time (Hypothesis 5). This provides an additional impediment for CDRs, because it implies that the more actors are required to agree, the sooner consensus cannot be reached anymore. Rewarding is therefore even more problematic to enforce than punishment, hence contributions are higher under punishment than reward.

The proportion of antisocial and cooperation-enforcing sanctioners in a population determines which decision rule leads to the highest contribution levels. Casari and Luini (2009) found, with relatively high levels of antisocial punishment, that sanctions on which two out of five actors agreed were much more effective than sanctions with an IDR. We use stricter collective decision rules of two and three out of four actors, in a population where antisocial sanctions hardly occur, and find that contributions are highest under an IDR. Future research could aim to extend these observations by assessing the effect of IDRs and CDRs in populations which are known to inhibit various levels of antisocial sanctioners.

We started this paper with the observation that many actors engaged in real-life public good problems use CDRs to successfully enforce cooperation. One possible reason that we find that an IDR is more effective, is that interactions in our experiment are one-shot and anonymous. In many real-life public good problems, especially in small communities or between nations, participants interact repeatedly. Moreover, actors can often communicate before deciding whether or not to sanction. Repeated interaction and communication imply that actors can coordinate on raising the required proportion of agreeing others. Also, often it is possible to identify which actors neglected to agree on sanctioning. Therefore, when the required consensus is not reached the actors who did not sanction can be held accountable, for example through second-order punishment (Cinyabuguma et al., 2006). Repeated interactions, communication on whom to sanction, and public announcement of sanctioning decisions can be implemented in future experiments to enhance resemblance with actual public good problems. These factors might explain why CDRs are often successful in real-life.

# LITERATURE CITED

Andreoni, J. (1988). Why free ride? strategies and learning in public goods experiments. *Journal of Public Economics 37*(3), 291–304.

Andreoni, J., W. Harbaugh, and L. Vesterlund (2003). The carrot or the stick: Rewards, punishments, and cooperation. *The American Economic Review 93*(3), 893–902.

Bochet, O., T. Page, and L. Putterman (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization 60*(1), 11–26.

Buchanan, J. and G. Tullock (1967). *The calculus of consent: Logical foundations of constitutional democracy*. Ann Arbor: University of Michigan Press.

Carpenter, J. (2007). The demand for punishment. *Journal of Economic Behavior & Organization 62*(4), 522–542.

Carpenter, J. and P. Matthews (2009). What norms trigger punishment? *Experimental Economics 12*(3), 272–288.

Casari, M. and L. Luini (2009). Cooperation under alternative punishment institutions: An experiment. *Journal of Economic Behavior & Organization 71*(2), 273–282.

Cinyabuguma, M., T. Page, and L. Putterman (2006). Can second-order punishment deter perverse punishment? *Experimental Economics 9*(3), 265–279.

Dari-Mattiacci, G. and G. De Geest (2010). Carrots, sticks, and the multiplication effect. *Journal of Law, Economics, and Organization 26*(2), 365–384.

Dawes, C., J. Fowler, T. Johnson, R. McElreath, and O. Smirnov (2007). Egalitarian motives in humans. *Nature 446*(7137), 794–796.

De Quervain, D., U. Fischbacher, V. Treyer, and M. Schellhammer (2004). The neural basis of altruistic punishment supporting online material. *Science 305*(5688), 1254–1258.

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

Denant-Boemont, L., D. Masclet, and C. Noussair (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory 33*(1), 145–167.

Diekmann, A. and T. Voss (2003). Social norms and reciprocity. *Arbeitsbericht des Instituts für Soziologie* (33).

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*(5), 495–511.

Fehr, E. and U. Fischbacher (2004). Social norms and human cooperation. *Trends in Cognitive Sciences 8*(4), 185–190.

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

Fehr, E. and H. Gintis (2007). Human motivation and social cooperation: Experimental and analytical foundations. *Annual Review of Sociology 33*, 43–64.

Fehr, E. and S. Gächter (2002). Altruistic punishment in humans. *Nature 415*(6868), 137–140.

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

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters 71*(3), 397–404.

Fischbacher, U. and S. Gächter (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review 100*(1), 541–556.

Greiner, B. (2004). The online recruitment system orsee 2.0-a guide for the organization of experiments in economics. *University of Cologne, Working paper series in economics 10*(23), 63–104.

Gächter, S. and B. Herrmann (2009). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences 364*(1518), 791–806.

Gächter, S. and C. Thöni (2011). Micromotives, microstructure, and macrobehavior: The case of voluntary cooperation. *Journal of Mathematical Sociology 35*(1-3), 26–65.

Herrmann, B., C. Thöni, and S. Gächter (2008). Antisocial punishment across societies. *Science 319*(5868), 1362–1367.

http://www.europa.eu (nov. 2010).

http://www.un.org (nov. 2010).

Kocher, M., P. Martinsson, and M. Visser (2008). Does stake size matter for cooperation and punishment? *Economics Letters 99*(3), 508–511.

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, 183–214.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. *Handbook of Experimental Economics, Princeton University Press, Princeton*, 111–194.

Masclet, D., C. Noussair, S. Tucker, and M. Villeval (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *The American Economic Review 93*(1), 366–380.

Milinski, M. and B. Rockenbach (2011). On the interaction of the stick and the carrot in social dilemmas. *Journal of Theoretical Biology*.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*, Volume 124. Harvard University Prress.

Ones, U. and L. Putterman (2007). The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *Journal of Economic Behavior & Organization 62*(4), 495–521.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Ostrom, E. (2000). Collective action and the evolution of social norms. *The Journal of Economic Perspectives 14*(3), 137–158.

Rand, D., A. Dreber, T. Ellingsen, D. Fudenberg, and M. Nowak (2009). Positive interactions promote public cooperation. *Science 325*(5945), 1272–1275.

Sefton, M., R. Shupp, and J. Walker (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry 45*(4), 671–690.

Selten, R. (1967). Die strategiemethode zur erforschung des eingeschräncht rationalen verhaltens in rahmen eines oligopolexperiments', in beiträge zur experimentellen wirtschaftforschung, ed. by h. sauermann, tübingen: Jcb mohr.

Veszteg, R. and E. Narhetali (2010). Public-good games and the balinese. *International Journal of Social Economics 37*(9), 660–675.

Walker, J. and M. Halloran (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics 7*(3), 235–247.

Wiedemann, V., D. Barrera, and V. Buskens (2012). Comparing consequences of carrots and sticks on cooperation in repeated public good games. *SSRN working paper*.