DISCUSSION PAPER SERIES

No. 5497

CHOOSING THE STICK OR THE CARROT? ENDOGENOUS INSTITUTIONAL CHOICE IN SOCIAL DILEMMA SITUATIONS

Matthias Sutter, Stefan Haigner
and Martin Kocher

INDUSTRIAL ORGANIZATION and PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP5497.asp**

ISSN 0265-8003

CHOOSING THE STICK OR THE CARROT? ENDOGENOUS INSTITUTIONAL CHOICE IN SOCIAL DILEMMA SITUATIONS

Matthias Sutter, University of Innsbruck and University of Cologne
Stefan Haigner, University of Innsbruck
Martin Kocher, University of Innsbruck and University of Amsterdam

Discussion Paper No. 5497
February 2006

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 5497

February 2006

ABSTRACT

Choosing the Stick or the Carrot? Endogenous Institutional Choice in Social Dilemma Situations*

We analyse an experimental public goods game in which group members can endogenously determine whether they want to supplement a standard voluntary contribution mechanism with the possibility of rewarding or punishing other group members. We find a large and positive effect of endogenous institutional choice on the level of cooperation in comparison to exogenously implemented institutions. This suggests that democratic participation rights enhance cooperation in groups. With endogenous choice, groups typically vote for the reward option, even though punishment is actually more effective in sustaining high levels of cooperation. Our results are evaluated against the predictions of social preference models.

JEL Classification: C72, C91 and C92
Keywords: endogenous institutional choice, experiment, public goods, punishment, reward and voting

Matthias Sutter
University of Cologne
Department of Economics
Albertus-Magnus-Platz
D-50923 Köln
Germany

Email: msutter@uni-koeln.de

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=149960

Stefan Haigner
University of Innsbruck
Department of Public Economics
Universitasstrasse 14
A-6020 Innsbruck
AUSTRIA

Email: stafan.haigner@uibk.ac.at

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=164024

Martin Kocher
University of Amsterdam
Department of Economics
CREED
Roetersstraat 11
NL-1018 WB Amsterdam
Netherlands
Email: martin.kocher@uibk.ac.at

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=162602

* We would like to thank Simon Gächter, Rudolf Kerschbamer, Michael Kosfeld, Stephan Kroll, Wolfgang Luhan, Nikos Nikiforakis, Axel Ockenfels, Louis Putterman and Christina Strassmair for very helpful comments and Olga Mellizo for language editing. We are grateful to the Center for Experimental Economics at the University of Innsbruck (sponsored by *Raiffeisen-Landesbank Tirol*) and the *Austrian National Bank* (Jubiläumsfondsprojekt No. 9879) for financial support. Kocher acknowledges financial support from the ENABLE Project under the European Union 6th Framework Program. The paper reflects the views of the authors, and the Community is not liable for any use that may be made of the information contained therein.

Submitted 16 January 2006

1 Introduction

In this paper we address the issue of cooperation in social dilemma situations using an experimental public goods game. In contrast to most previous studies, we let subjects decide endogenously on the institutions governing their interactions. In particular, subjects can vote on whether they want to supplement a standard voluntary contribution mechanism (VCM) with the possibility of rewarding or punishing other subjects or whether they prefer to use the standard VCM. We find that the endogenous institutional choice has a large and positive effect on cooperation levels in the public goods game compared to a setting in which these institutions have been determined exogenously (by the experimenter). This suggests that democratic participation rights foster cooperation, and the effect is in addition to the previously observed cooperation-enhancing effect of reward and punishment (when these institutions have been implemented exogenously).

By varying the effectiveness of the available punishment or reward technology (defined as the ratio of monetary consequences of being punished or rewarded to the costs of punishing or rewarding), we are able to show that the reward option is chosen almost exclusively when reward or punishment are highly effective (i.e. have a high ratio), whereas the standard VCM is the preferred institution when the technology is less effective. The punishment option – though very effective in raising contributions – is only very rarely chosen endogenously, indicating that most subjects prefer an institutional environment without punishment.

We compare our results to predictions based on standard game theoretic analysis (under the assumptions of rationality and selfishness) and predictions derived from a model considering inequity aversion (Fehr and Schmidt, 1999). While the standard game theoretic model essentially fails to explain contributions and institutional choices, the model with social preferences is able to predict many, but not all behavioral regularities from our experiment. In particular, the inequity aversion model fails to account for the observed behavior under the reward institution and for the frequency with which subjects participate in the costly vote on the institution.

Our paper is related to a large number of studies dealing with the conditions and institutions that foster cooperation in social dilemma situations. In recent years, the effects of reward and especially of punishment on cooperation in social dilemma games have received considerable attention, for instance in Fehr and Gächter (2000, 2002), Sefton et al. (2002), Andreoni et al. (2003) or Masclet et al. (2003). These and most subsequent papers

mentioned below share the feature that the institution governing the interaction of subjects has been determined exogenously by the experimenters. Though these studies have provided illuminating and very important insights into the determinants of cooperation in groups, they are unable to answer the two straightforward questions that are central to our paper: (1) Which institutions do subjects actually prefer and choose themselves in a social dilemma? (2) How does the endogenous choice of rules affect group members' interaction and their level of cooperation? The answers to both questions are important in order to judge whether the inferences about the treatment effects of different – but exogenously imposed – institutions are reliable and can be generalized to situations where subjects can self-select or avoid certain institutions through an endogenous choice.

In real life, groups often establish themselves the (explicit or implicit) rules that govern their interaction. On a large-scale international level, for instance, one might think of countries' representatives convening in Kyoto and the successor host cities of the United Nations Framework Convention on Climate Change. These representatives establish rules – and possible sanctions in case of their violation – on how to provide the public good of carbon dioxide emission reduction. Regarding small group interactions, work teams in companies typically develop social norms on how to deal with 'rate busters' or on how to reward team members for contributions to the team success. In the private sphere, people sharing a flat decide on whether they want to impose rewards or sanctions for the private contributions to public goods like cleaning, shopping, cooking, disposing garbage, and so on. Obviously, reward and punishment do not always exist as a monetary fine or a monetary gift, but also in the form of social approval or disapproval or in many other forms that might be at least as effective.[1]

Another example of the endogenous determination of rules is the use of local common pools (like a fish-pond or a pasture for cows) in medieval times. Villagers could typically vote on how to use the common pool and on which sanctions to apply in cases of misconduct (see Ostrom et al., 1992; Casari and Plott, 2001). In fact, the paper by Casari and Plott (2001) is an experimental test of an institution that had governed the use of a common pool resource in the Italian Alps for centuries. The institution was developed

[1] One of the authors vividly remembers that while sharing a flat in college, the following rule was jointly implemented and strictly enforced by social pressure: The resident who most often failed to reach the deadline for cleaning the bathroom and the kitchen according to a pre-arranged cleaning plan was obliged to invite the other two residents out to dinner in a restaurant of their choice.

endogenously by every village through a collective choice. It specified not only the use of the common pool resource, but also in advance a monitoring and sanctioning system and the amount of the fine for a violation of the rule. In the experiment of Casari and Plott (2001), though, the institution had been exogenously fixed by the experimenters, i.e. there was no endogenous choice.

As already briefly mentioned, we build upon a large and rapidly growing body of studies on the effects of *exogenously* imposed reward and punishment options in social dilemmas. Fehr and Gächter (2000, 2002) have analyzed the contributions to a public good when subjects can punish each other after having observed the other group members' contributions. Even though using the punishment option is at odds with selfish money-maximization in their experimental design, it is used rather frequently, thereby inducing very high levels of cooperation.[2] In order to study the motives behind sanctions, Falk et al. (2005) have designed a prisoner's dilemma experiment with a punishment option. By imposing different costs on the punisher in different treatments, they are able to analyze the driving forces of punishment and whether and to which extent these forces can be explained by models that incorporate social preferences. Their main finding is that cooperators tend to punish defectors even if the punishment technology does not allow them to reduce the payoff differential. Furthermore, Falk et al. (2005) observe a considerable amount of spiteful punishment, where defectors punish cooperators. This occurs, however, almost exclusively in settings in which punishment is effective.

A very important feature in an analysis of the impact of punishment is the demand for punishment under varying prices and/or effectiveness. While some papers use only one punishment technology, several other papers, including ours, compare the behavioral impact of differently effective punishment (and reward) mechanisms. Only a few papers concentrate on eliciting a complete demand function for punishment by systematic variations of prices of punishment. Two examples are Anderson and Putterman (2006) as

[2] In political science or social psychology, related results have been documented earlier by Yamagishi (1986) or Ostrom et al. (1992), for example. Recently, the emerging field of neuro-economics has examined the physiological foundations of punishment. de Quervain et al. (2004) have shown that using the punishment option actually triggers some kind of reward for the punishing subject, as it stimulates the very area in the brain that is also activated when one is being rewarded. This finding is one rationalization of the fact that punishment is even used in one-shot interactions (respectively in the last round of repeated interactions) or when only third parties that are materially unaffected by the actions of the players can punish (Fehr and Fischbacher, 2004).

well as Carpenter (2006). Both papers essentially show that the law of demand holds: the quantity of punishment demanded is lower with a higher price.

Masclet et al. (2003) have addressed the question of whether the negative monetary consequences of punishment strengthen cooperation or whether the expression of disapproval of another subject's behavior through punishment is possibly equally or even more important. They have compared a public goods game with monetary sanctions to one with only symbolic (non-monetary) sanctions. Cooperation remains higher with the monetary sanctions, but even the non-monetary sanctions raise cooperation levels above the level of the standard VCM. Rege and Telle (2004) also report a positive effect of (non-monetary) social (dis)approval on cooperation in a public goods game. Contributions to the public good increased significantly when subjects had to publicly indicate their contributions, compared to a situation where subjects remained completely anonymous.

Kroll et al. (2005) study the role of (monetary) punishment in yet another context where group members can vote on non-binding minimum contributions to a public good. When failure to comply with the non-binding minimum can be punished, contributions increase significantly. Thus, it is not the outcome of the non-binding vote that is important for higher contributions in such a setting, but the possibility of punishing defectors.

One of the few papers that take reward into account is Sefton et al. (2002). They have included a reward option in their public goods experiments. Like punishment, reward has been found to increase cooperation in comparison to the standard VCM-setting. Reward, however, seems to be less suitable for sustaining high levels of cooperation, as the decay of contributions over time is faster with the reward option than with the punishment option. Hence, using the carrot (i.e., rewards) seems to be less efficient for enhancing cooperation in a group than the stick (i.e., punishment). A similar result is obtained by Fehr and Gächter (2003). They find punishment to be more effective than reward in mitigating the free-rider problem and also observe a strong impact of the effectiveness of reward and punishment. Andreoni et al. (2003) provide complementary results in a kind of bargaining game where a proposer can offer a portion of a pie to a responder, and the responder can reward and/or sanction the proposer for her offer. It turns out that the distribution of offers is much more affected by the punishment option than the reward option, even though the latter also has an effect.[3]

[3] Reward in another interactive context – in the form of (voluntary) bonus contracts in a principal-agent relationship – is studied, e.g., by Fehr et al. (2001) or Fehr and Schmidt (2004).

Public goods games with endogenous institutional choice have only very recently received scholarly attention. Botelho et al. (2005) extend the design of Fehr and Gächter (2000, 2002) by allowing subjects to choose directly whether they want to operate under a punishment institution or under a standard VCM. After subjects have experienced both the VCM and the VCM with the punishment option in a (partly perfect) stranger design, they can choose the governing institution for one final round (with substantially increased salience). Using the Fehr and Gächter (2000) punishment technology, Botelho et al. (2005) find that almost none of their cohorts opt for the punishment institution. From an ex-post point of view, this choice is justified because the overall efficiency of interactions (including punishment costs) is significantly lower in a world with punishment than without. Compared to Botelho et al. (2005), our experiment considers not only a punishment, but also a reward institution, which will be shown to be rather attractive among groups when it is compared to the punishment institution. Furthermore, we use a partner design because we are mainly interested in the institutional choice of groups whose members will interact repeatedly after the initial decision on the governing institution. Finally, in contrast to Botelho et al. (2005) our subjects may abstain from participating in the vote. We are, therefore, able to draw inferences on the relationship between vote participation and behavior in the public goods game.

Gürerk et al. (2005) study the *self-selection of individuals* into groups with an *exogenously given* institution (like reward or punishment), and they find that subjects self-select into groups with punishment (when the standard VCM is the alternative) much more often than into groups with rewards (also with the standard VCM as alternative). Our approach is different in several aspects: First, we let *groups self-select the institution* that shall apply to their interaction. Therefore, our approach resembles a constitutional choice to a much greater extent than the voting by feet-approach of Gürerk et al. (2005). Second, we compare the attractiveness of reward and punishment when they are equally effective and when subjects can choose directly between both institutions. In Gürerk et al. (2005) punishment had much stronger consequences than reward (which may explain why subjects preferred punishment over reward in their experiment) Third, in Gürerk et al. (2005) subjects could switch between different institutions as often as they wished to without any costs. In our design, we implement small costs of voting in order to be able to assess potential effects of vote participation on cooperation, and sanctioning and rewarding

behavior. Group members are not required to participate in this vote, but those abstaining are bound by the collective decision of voters in their group.

A straightforward and related question concerns the endogenous choice of a punishment and/or reward technology. So far, there is only evidence on endogenously chosen punishment technologies. Ertan et al. (2005) show that most groups rule out the possibility of punishing high contributors, but most groups implement an institution where the punishment of low contributors is possible. An earlier paper with a similar focus is Decker et al. (2003). They study the effects of different voting rules – like unanimity or simple majority – on cooperation. In their design, group members could propose their desired punishment intensity for each group member, and in one particular treatment, they could determine the most preferred voting rule by submitting bids in a first-price auction. The results of Decker et al. (2003) suggest that the unanimity rule is the most preferred one. Contrary to our paper, none of these studies had a control treatment with exogenously imposed institutions. Consequently, they have not been able to address the effects of endogenous institutional choice on cooperation.[4]

Our voting procedure links our paper to the more general question of whether democratic participation rights have an influence on the behavior of subjects and, consequently, how they influence individual behavior. Frey (1994), Frey et al. (2004) or Pommerehne and Weck-Hannemann (1996), for instance, have studied the effects of the degree of direct-democratic participation rights (as a form of endogenous institutional choice) on the level of cooperation among citizens and between citizens and the public sector. Their findings indicate that stronger direct-democratic participation rights increase cooperation and reduce, for example, tax evasion (see also Feld and Tyran, 2002). As already mentioned, in our experiment we examine how the decision to participate in the vote on the institution is linked to a subject's contribution level in the public goods game and to her punishing and rewarding behavior. Hence, we can establish a link between

[4] Of course, the experimental literature on endogenous institutional choice is much larger when one looks beyond public goods games. Kirchsteiger et al. (2005), for instance, investigate which information standards evolve endogenously when labor market participants can decide which other market participants to inform about own trade offers. Brown et al. (2004) provide evidence for the endogenous emergence of long-term relationships between firms and workers in the absence of enforcement of contracts by a third party. Other labor market experiments with endogenous institutions (for example on the payoff scheme) are Eriksson and Villeval (2004), Eriksson et al. (2005) or Lazear et al. (2005).

voting and economic decision making. We are also able to measure the economic benefit of an endogenous institutional choice by comparing contributions under the endogenous choice with contributions under identical, but exogenously determined, institutions (a comparison that is hardly possible with field data). Our results provide clear evidence that the benefit of an endogenous institutional choice is positive. In other words, we observe a clear democratic participation rights-premium to cooperation.

The remainder of this paper is structured as follows. In Section 2 we present the public goods game and the different institutions that may be implemented. Section 3 describes the experimental design, and in Section 4 we derive several hypotheses on contribution levels and on voting behavior. Section 5 reports our experimental results, and finally Section 6 discusses our findings and concludes the paper.

2 The public goods games

Let $I = \{1, 2, \ldots, n\}$ denote a group of n subjects who interact in T periods. In each period $t \in \{1, 2, ..., T\}$ individual $i \in I$ receives an endowment E which can be allocated either to a private good or a public good. The voluntary contribution of individual i to the public good in period t, $c_{i,t}$, must satisfy $0 \le c_{i,t} \le E$. Let C_t denote the sum of all group members' contributions (i.e. $C_t = \sum_{j=1}^{n} c_{j,t}$). Individual member i's payoff from her contribution in period t is given by

$$\pi_{i,t} = E - c_{i,t} + \gamma C_t \tag{1}$$

The marginal per capita return (MPCR) from investing into the public good is denoted as γ, which satisfies $0 < \gamma < 1 < n\gamma$.

We consider three variants of this game: (*i*) the standard voluntary contribution mechanism (VCM) as it has just been described, (*ii*) the VCM with a punishment option, and (*iii*) the VCM with a reward option. In the standard VCM, all decisions are made simultaneously and each period has only one stage (the contribution stage). In the VCM with punishment (reward) each period has two stages, first the contribution stage and, thereafter, a punishment (reward) stage. In the second stage, group members receive information about the other members' contributions and can punish (reward) each of the

other group members. Punishment is costly for the punisher and the punished subject. Reward is costly for the rewarding subject and beneficial for the rewarded subject.

We implement reward or punishment as a *binary* decision only. Hence, a group member cannot choose between different intensities of reward or punishment. The effectiveness of reward or punishment, however, is exogenously varied. We use the term leverage and the symbol L to refer to different technologies. We consider $|L| = l$ or $|L| = h$, with $h > l$. A higher leverage means that the same choice of punishing or rewarding another subject has larger monetary consequences for the punished or rewarded subject (without increasing the costs for the subject who punishes or rewards).

Taking into account the monetary consequences of the second stage in each period yields the following payoff function for member i in the VCM with punishment, or reward.

$$\pi_{i,t} = E - c_{i,t} + \gamma C_t + L\sum_{k \neq i} p_{ik,t} - \sum_{h \neq i} p_{hi,t} \tag{2}$$

where $p_{hk,t} = 1$ if member k has punished (rewarded) member h in period t and zero otherwise.[5] The sign of the parameter L indicates whether we are in the game with punishment ($L < 0$) or the game with reward ($L > 0$). The absolute value of L captures the effectiveness of the punishment (reward) technology, with a larger L indicating a stronger effect (a higher leverage) of one's own actions on the profit of others.[6]

3 Experimental design and procedure

Our experimental treatments were based on the public goods games introduced in the previous section. The parameters were set up as follows: group size $n = 4$, endowment per

[5] Normalizing the cost of punishment (reward) is always possible by an appropriate choice of L. Note that if the last two terms in equation (2) are dropped (or the punishment and reward option are not used), we are back in the standard VCM given by equation (1).

[6] Needless to say, it is impossible to design a straightforward mechanism where punishment and reward are equivalent in terms of efficiency, unless the Nash equilibrium (assuming selfishness and common knowledge of rationality) is played (see Section 4 for a more detailed discussion). By nature, costly punishment makes everybody worse off if used, and costly reward has the potential of increasing overall efficiency (if $|L| > 1$). Since all real-world sanction and reward mechanisms have this feature, it seems natural to preserve it when studying the choice between reward and punishment in the laboratory.

period $E = 20$, MPCR $\gamma = 0.4$, and number of periods $T = 10$. We used a 3×2×2 factorial design with the following three factors:

F1. The institution governing the provision of the public good: either (*i*) the *standard VCM*, (*ii*) the VCM with *punishment*, or (*iii*) the VCM with *reward*.

F2. The leverage of punishment or reward: either (*i*) $|L| = 1$ or (*ii*) $|L| = 3$.[7]

F3. The way in which the institution is determined: either (*i*) *exogenously* by the experimenter or (*ii*) *endogenously* by the group itself through a voting procedure.

The first two factors of our design were explained in the previous section. The way in which the institution is determined (factor F3) requires some additional explanation. In the exogenous treatments, the institutions were simply pre-determined by the experimenter. Participants in these sessions were informed in the experimental instructions about the institution under which they were playing the public goods game.

In the endogenous treatments, participants were introduced to all three possible institutions and could determine the institution prevailing in their group by the following procedure: First, subjects had to decide whether they wanted to participate in a costly vote on the institution. Voters incurred a small one-time fee of 10 ECU (experimental currency units). Non-voters did not incur any costs, but were aware that the decision of voters would be binding for them. Then, the following approval voting protocol with simultaneous choices of all voters was applied: Each voter had to indicate for each available institution (*standard-VCM*, *punishment*, or *reward*) whether she was willing to accept this institution or not.[8] The voting outcome was determined as follows: If unanimous consent for any of the three institutions was not reached in the first ballot, the vote was repeated until

[7] Though one can find a wide range of technologies for the *punishment* option (ranging from a 1:1 ratio in Sefton et al., 2002, to a 1:5 ratio in Andreoni et al., 2003), the comparative statics of different technologies on contribution levels have only recently been examined systematically. See Masclet and Villeval (2004), Anderson and Putterman (2006), Casari and Luini (2005), Egas and Riedl (2005), Nikiforakis and Normann (2005) or the survey in Casari (2005) besides our own contribution. The only paper that we are aware of that is attempting to analyze different *reward* technologies in a very similar setting to ours is Fehr and Gächter (2003).

[8] It is important to note here that the endogenous choice was always restricted to a given leverage $|L|$. Hence, in one endogenous experimental treatment subjects could choose between the standard VCM, punishment with $|L| = 1$, or reward with $|L| = 1$. In the other endogenous experimental treatment, the leverage was always $|L| = 3$.

unanimity (for at least one institution) was reached.[9] In the case of a unanimous consent of all voters for exactly one institution, this institution was applied in the public goods game for all ten periods. If more than one institution had received unanimous support, then the institution was determined by a random draw from the tied institutions.[10] Finally, if none of the four group members had participated in the vote, then the institution was determined randomly in order to avoid creating a bias by imposing a default institution.

Voters were always informed about the total number of voters in their group, and after each voting round they were also informed about the number of votes each institution had received. Voting decisions were secret in the sense that they could not be linked to individuals or individual behavior in the consecutive public goods game.

One noteworthy feature of our design is that there was not a history of the VCM before participants decided on whether to vote and what to vote for. While one might claim that this is a disadvantage of our design – because subjects had to make an institutional choice before having experienced a particular institution – we believe it is rather a strength, given that we are mainly interested in the kind of rules that subjects prefer when facing a social dilemma situation. As such, our focus is on the constitutional choice of rules, and not on how the choice of rules depends on previous experience with different rules.[11]

---

[9] There was no upper limit on the number of voting rounds. In the results section we will show that most groups reached a consensus rather soon. Even the groups with the highest number of voting rounds before reaching an agreement finished the voting procedure in less than ten minutes (because the software yielded instantaneous feedback after the last voter in a group had confirmed her voting decisions). Note that the costs of participating in the vote were independent of the number of voting rounds, and therefore had to be paid by voters only once.

[10] Of course, we do not argue that our voting mechanism is actually applied to real-world institutions for social dilemmas. We are confident, however, that its structure comes close to more or less tacit agreements on the implementation of sanctions and rewards in many small groups in real life and also to the decision making in institutional arrangements in larger international organizations like in the context of the Kyoto protocol or World Trade Organization negotiation rounds (where unanimity is often required to implement an institution).

[11] In order to check whether more experience with the game could lead to different results regarding the endogenous choice, we asked subjects in a post-experimental questionnaire whether they would stick to their institutional decision if the experiment had continued or whether they would now support another institution. The results from this questionnaire reinforce our conclusions from the actual institutional choice and are presented in detail in Section 5.2.1.

We used a partner matching in the experiment, which yields one statistically independent observation per group of four subjects. In the punishment or reward stage that followed the contribution stage, subjects were informed about the contributions of all of their group members, and could decide on which members to punish or reward (depending on the prevailing institution).[12] It was possible to punish or reward any other member and even more than one group member. At the end of each period subjects were again informed about the other members' contributions, and about the decisions on punishment and reward (if applicable) and the resulting payoffs.

In the exogenous treatments we had in total 200 participants, yielding ten independent observations for each of the five different exogenous treatments (i.e. VCM, punishment with $L = -1$ or $L = -3$, and reward with $L = 1$ or $L = 3$). In the endogenous treatment we let 160 participants choose their desired institution when the leverage of reward or punishment was fixed at $|L| = 1$. Hence, we have 40 independent observations for the institutional choice with the low leverage. In addition, we had another 80 participants (i.e. 20 independent groups) choosing the institution under the high leverage, i.e. with $|L| = 3$.

In sum, 440 students drawn from the general undergraduate population at the University of Innsbruck participated in the computerized experiment (using z-Tree by Fischbacher, 1999). No subject was allowed to participate in more than one session and none of them had participated in a public goods experiment before.[13] Decisions were taken anonymously in cubicles and communication among participants was not permitted. The average session length was less than one hour, and subjects earned about 10 € on average. Table 1 summarizes our experimental treatments and the number of independent observations (i.e. groups) in each treatment.

---

[12] Each member of a group had a unique ID. Contributions to the public good could be assigned to this ID throughout the experiment. Decisions on rewards and punishment, however, could not be linked to the ID. Hence, subjects could not identify which other group member had punished or rewarded them. This was done so in order to avoid retaliation (of one member punishing another for having been punished by the other member in the past). Such "counter-punishment" has been shown to occur, for instance, by Denant-Boemont et al. (2005) as well as Nikiforakis (2005).

[13] The experimental instructions for endogenous choice under $|L| = 3$ can be found in Appendix A. The instructions for the other treatments are analogous and available upon request. All instructions were context-free and neutrally framed.

Table 1 about here

4 Predictions

Before moving on to the experimental results, we would like to formulate two sets of predictions about the behavior in the different treatments. We start with straightforward predictions based on the assumptions of purely selfish and rational decision makers and move on to predictions derived from a model that takes social preferences into account. For both approaches we resort to the assumptions of common knowledge of rationality and risk neutrality throughout the remainder of the paper. In order to be succinct, only the main predictions and an intuition for each will be presented in the text. The interested reader is referred to Appendix B for the formal derivations and proofs.

4.1 Predictions when agents are purely selfish

Assuming that subjects care only for their own monetary payoffs, it is obvious that a rational, money-maximizing agent shall abstain from costly punishment or reward. Since the game is finitely repeated and we assume common knowledge of rationality, backward induction yields the well-known result that contributions do not depend on the availability of punishment or reward, irrespective of their leverage. Consequently, the standard VCM, the one with punishment and the one with reward will all yield the same level of contributions, which is zero (i.e. $c_{i,t} = 0$ for all t) because free-riding is a dominant strategy due to $\gamma < 1$.

Prediction 1*: Contributing zero is the only equilibrium under all available institutions.*

Since contributions are not expected to differ, then subjects should be indifferent between the possible institutions and should, thus, abstain from voting in order to save the voting costs.

Prediction 2*: In the endogenous treatments, subjects do not participate in the costly vote because they are indifferent between the possible institutions.*

4.2 Taking social preferences into account

Since models that incorporate fairness concerns into the utility function of subjects have been very successful in explaining behavior in public goods games with or without punishment, it seems warranted to strive for alternative predictions based on such models. Of course, there are several models that could be applied in our context. We decided to rely on the influential model by Fehr and Schmidt (1999), henceforth abbreviated by FS. One reason for choosing the FS-model is its assumption that subjects compare their own material payoff with the material payoff of each other member in their reference group.[14]

FS assume that the utility of subject i in a set of n subjects depends on the vector of monetary payoffs $x = (x_1, \ldots, x_n)$ in the following form:

$$U_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_i - x_j, 0\} \qquad (3)$$

with $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$. The second term on the right-hand side of equation (3) captures the utility losses from disadvantageous inequality, and the third term the losses from advantageous inequality. Purely selfish subjects are affected by neither since for them it is assumed that $\alpha_i = \beta_i = 0$.

4.2.1 Contributions in the standard VCM

FS show (in their Proposition 4 on p. 839) for the standard VCM that

(a) If $\gamma + \beta_i < 1$ for member i, then it is a dominant strategy for that member to free-ride and choose $c_{i,t} = 0$ in all periods.

(b) Let f denote the number of members with $\gamma + \beta_i < 1$, where $0 \leq f \leq n$. If $f/(n\text{-}1) > \gamma/2$, then there is a unique equilibrium with $c_{i,t} = 0$ for all members i.

[14] Hence, a model of social preferences where subjects compare their own payoff only to the *average* payoff in their group (as in another influential model by Bolton and Ockenfels, 2000, for instance) seems less straightforward to be used in our context. Falk et al. (2005) have shown, in a similar setting like ours, that the FS-model is more suitable in such a context. Of course, we acknowledge that there are many other fairness models that may be relevant here, and it is obviously impossible to apply all of them to our experiment. As we discuss in the conclusion, intention-based models might prove helpful in explaining several of our stylized facts. These models are, however, inappropriate for deriving clear predictions, especially for our endogenous institutional choice.

(c) Only if $f/(n\text{-}1) < (\gamma + \beta_j - 1)/(\alpha_j + \beta_j)$ for all members j with $\gamma + \beta_j \geq 1$ (the latter being called '*cooperators*' in the following), there is a multiplicity of equilibria where members with $\gamma + \beta_i < 1$ contribute zero and the others contribute $c_{i,t} = c \in [0, E]$. Under this condition it must hold that $(\gamma + \beta_j - 1)/(\alpha_j + \beta_j) < \gamma/2$.

For the intuition and the formal proofs we refer the reader to Proposition 4 of Fehr and Schmidt (1999, p. 839 and pp. 860-863). Note that from part (b) it follows for our case with $\gamma = 0.4$ and $n = 4$ that if $f/(n\text{-}1) > 0.2$ (i.e. the fraction of members with a $\beta_i < 0.6$ is larger than 20%), then there is no equilibrium with positive contributions in the standard VCM. This means that one member with $\gamma + \beta_i < 1$ suffices to wipe out cooperation in the whole group.

Proposition 1*: Equilibrium contributions in the standard linear VCM depend on the distribution of social preferences in a group. If at least one member cares relatively little about advantageous inequality (i.e., $\beta_i < 0.6$), the only equilibrium is complete free-riding ($c_{i,t} = 0$) of all members. Otherwise, there is a multiplicity of equilibria with all members contributing $c_{i,t} = c \in [0, E]$.*

4.2.2 Punishment or reward with |L| = 1

In order to consider the case of punishment under the low leverage $|L| = 1$, note that free-riding creates a material payoff advantage for the free-rider relative to those who cooperate. Let us denote the number of cooperators (who satisfy $\gamma + \beta_i \geq 1$) by n'. As long as $|L| \leq 1$, cooperators cannot reduce their payoff disadvantage by punishing free riders. As a consequence, as long as $n' < n$ there is no equilibrium strategy in which punishment is a credible threat. Consequently, group members face the same incentives as in the standard VCM. Since $n' < n$ implies that there is at least one group member with $\beta_i < 0.6$, then full free-riding is the only equilibrium. Only if *all* group members are cooperators (hence, $n' = n$) then there is an equilibrium where punishment would be credible but would never be used because all members are cooperators anyway. In the latter case, all contribution levels $c_{i,t} = c \in [0, E]$ can be maintained as an equilibrium.

***Proposition 2**: Consider the VCM with punishment and $|L| = 1$. Complete free-riding is the only equilibrium when at least one group member satisfies $\beta_i < 0.6$ (which yields the same prediction as in the VCM). Only if all members satisfy $\beta_i \geq 0.6$, any positive contribution can be sustained as an equilibrium.*[15]

For the reward institution under $|L| = 1$ there is no equilibrium where the reward option is used as long as the number of cooperators (with $\gamma + \beta_i \geq 1$) is strictly smaller than the number of group members. Hence, for $n' < n$ we arrive at the same equilibrium predictions as in the standard VCM. Only if $n' = n$ and $\beta_i \geq 0.75$ for all members, can mutual rewards be part of an equilibrium strategy where contributions are $c_{i,t} = c \in [0, E]$ like in the standard VCM.

***Proposition 3**: Consider the VCM with reward and $|L| = 1$. If at least one group member satisfies $\beta_i < 0.6$, zero contributions are the only equilibrium. Only when all members are rather averse to advantageous inequality ($\beta_i \geq 0.6$ for all i) then positive contributions $c_{i,t} = c \in [0, E]$ are feasible in equilibrium. Using the reward option can only be part of an equilibrium strategy if all members satisfy $\beta_i \geq 0.75$.*[16]

If we disregard the unlikely case of $n' = n$, our predictions 4 and 5 imply the following consequence for the voting stage of the endogenous treatment under $|L| = 1$.

***Proposition 4**: Suppose group members can vote between the standard VCM, the VCM with punishment, and the VCM with reward under $|L| = 1$. If voting is costly, subjects do not participate in the vote because contributions are expected to be the same for the standard VCM and reward or punishment with $|L| = 1$.*

---

[15] This latter case is rather unlikely. If we take, for the purposes of illustration, the assumptions about the distribution of social preference parameters as reported in Fehr and Schmidt (1999, p. 844), there will be 60% of subjects with $\beta_i < 0.6$. The latter implies that only in about 2.6% of all cases will *all* group members satisfy $\beta_i \geq 0.6$ (if $n = 4$) and that in the remaining 97.4% of cases the only equilibrium will be zero contributions (because at least one member satisfies $\beta_i < 0.6$).

[16] Obviously, using the reward strategy with $|L| \leq 1$ does not increase overall efficiency in terms of payoffs.

4.2.3 Punishment or reward with |L| = 3

Under the punishment institution with $|L| = 3$, cooperators with positive contributions can increase their utility by actually punishing free-riders if cooperators suffer relatively strong from disadvantageous inequality (i.e. if they have a rather high α_i). When the threat of using the punishment option is credible, positive contributions of *all* group members – including potential free-riders – may be sustained as an equilibrium.

Proposition 5*: Consider the following strategy in the VCM with punishment and |L| = 3. In the first stage of each period each member contributes $c \in [0,E]$. If all members do so, there is no punishment in the second stage. This constitutes a subgame-perfect equilibrium if the following two conditions are satisfied:*

(i) $L \geq \frac{(1-\gamma)(c-c_{i,t})}{n'}$, *with $c_{i,t} < c$ denoting a free-rider's contribution, and*

(ii) $\frac{L}{k} \geq (n-n') + \frac{1}{\alpha_i}\left[(n-1) - \beta_i(n'-1)\right]$ *for all n' cooperators.*

This strategy constitutes a continuum of equilibria in the game where $c \leq \bar{c} = Ln'/(1-\gamma)$ can be enforced, but where no punishment actually occurs. Consequently, contributions are expected to be higher with punishment under |L| = 3 than under |L| = 1 and higher than in the standard VCM.

Condition (*i*) takes care of the incentives of a potential free-rider and ensures that the utility from contributing c is at least as high as the utility from contributing $c_{i,t} < c$. In line with basic intuition, this condition shows that the higher the marginal returns from the public good (γ) and the higher the number of cooperators (n'), the lower the punishment efficiency $|L|$ needs to be to sustain a free-rider's incentive to cooperate. Condition (*i*), however, shows that for a given $|L|$ there is an upper bound for the contribution c that can be sustained as an equilibrium. Let us denote this upper bound as $\bar{c} = Ln'/(1-\gamma)$ in the following, assuming that $c_{i,t} = 0$. For $n' = 3$ we get $\bar{c} = 15 < E$, for instance. Hence, full contributions cannot be enforced in our experimental design, unless all members satisfy $\gamma + \beta_i \geq 1$.[17]

[17] The reason for the existence of this upper bound is our weaker punishment technology as compared to Fehr and Gächter (2000), for instance. Whereas their design allowed for different intensities of punishment –

Condition (*ii*) is critical for the threat of punishment to be credible. It requires that cooperators benefit from punishing. The gains from punishment through a decrease in inequality must outweigh the costs of punishment – where the latter cover the monetary costs of punishment, but also possibly include the costs from increasing inequality relative to the other cooperators. Besides having an incentive to punish free-riders, cooperators must also not deviate from contributing c themselves. This is satisfied due to $\gamma + \beta_i \geq 1$.

For the reward institution under $|L| = 3$, we get the same prediction as under $|L| = 1$. As soon as one member satisfies $\gamma + \beta_i < 1$, the only equilibrium implies zero contributions of all members and no rewards. The equilibria where all members reward each other (if $n' = n$ and $\beta_i \geq 0.6$ for all members) are, of course, much more efficient than under $|L| = 1$, because rewarding each other under $|L| = 3$ creates an overall efficiency gain. Therefore, in these (unlikely) cases the reward option should become more appealing from a behavioral perspective. Note, however, that it does not change the set of feasible contributions $c_{i,t} = c \in [0,E]$ in equilibrium.

Proposition 6*: Consider the VCM with reward and $|L| = 3$. Equilibrium predictions in this institution are identical to those with reward under $|L| = 1$. Yet, equilibria where all members reward each other are more likely than under $|L| = 1$ because they require $\beta_i \geq 0.6$ instead of $\beta_i \geq 0.75$ for all members. Furthermore, mutual reward is much more efficient with $|L| = 3$ than with $|L| = 1$. Hence, more rewarding should take place under $|L| = 3$ than under $|L| = 1$.*

Concerning voting behavior in our endogenous treatment under $|L| = 3$, it is important to note first that the punishment institution is the only one in which a contribution level $\bar{c}$ can actually be enforced (with the magnitude of $\bar{c}$ depending on the number of cooperators in the group). In the standard VCM or with reward, positive contributions may be sustained in equilibrium, but only if none of the group members satisfies $\beta_i < 0.6$. Recall the remark in footnote 15 regarding the probability that one group member has $\beta_i < 0.6$, given a certain distribution of social preferences taken from FS. In

which made it possible to equate both cooperators' and free-riders' final payoffs for any possible difference after the contribution stage – our punishment technology is binary and has only a limited scope for reducing a free-rider's payoff.

97.4% of all cases the only equilibrium would be zero contributions (because at least one member has $\beta_i < 0.6$). In the punishment institution, though, it suffices to have one subject with $\beta_i \geq 0.6$ to be able to enforce a contribution of $\bar{c} = 5$. Given the distributional assumptions from FS, about 87% of groups should have at least one subject with $\beta_i \geq 0.6$. It can be easily verified that for almost any meaningful assumption about the distribution of social preferences among subjects, it is optimal (in expected terms) to participate in the vote (and incur the one-time cost of 10 ECU) and support only the punishment institution.

Proposition 7*: Suppose group members can vote between the standard VCM, the VCM with punishment, and the VCM with reward under |L| = 3. Assume that the distribution of social preferences follows FS as described above. Then, subjects are expected to support only the punishment institution and voter turnout will be higher under |L| = 3 than under |L| = 1.*

5 Experimental results

5.1 Behavior with exogenously determined institutions

Figure 1 displays the time trend of contributions over the ten periods in the different treatments with exogenous institutions, and panel (A) of Table 2 reports the overall averages. Contributions are lowest in the standard VCM-treatment, where subjects contribute on average 32.5% of their endowment (= 6.50 out of 20). This is clearly above the zero contribution level expected under the standard approach of common knowledge of rationality and selfishness (see Prediction 1), but it is not ruled out when taking into account social preferences (see Proposition 1). With the low leverage $|L| = 1$, the average contributions are only slightly higher with punishment (8.79) or with reward (6.62) than in the standard VCM. None of the pairwise comparisons of contributions yields a significant difference, though. This is in accordance with Propositions 2 and 3. From panel (B) of Table 2 it is also easy to see that profits are not significantly different between the standard VCM and the VCM with punishment or reward with $|L| = 1$.

Result 1. *When the institution is fixed exogenously and when |L| = 1, the institution does not influence the level of contributions. Hence, reward and punishment options do not lead to higher contributions than the standard VCM when their effectiveness is low.*

Table 2 and Figure 1 about here

When we look at contribution levels under the high leverage $|L| = 3$, we find significantly higher contributions with punishment (12.11) or reward (11.20) than in the standard VCM (6.50) ($p < 0.05$ for both pairwise comparisons; two-sided Mann-Whitney U-test[18]; $N = 20$ in each case). We find no significant difference in contribution levels between the punishment and reward treatments, though.

When we examine the impact of the leverage for a given institution, i.e., the effectiveness of the mechanisms, we find significantly higher contributions under the high leverage ($|L| = 3$) than under the low leverage ($|L| = 1$) both for punishment (12.11 vs. 8.79; $p < 0.05$; Mann-Whitney U-test, $N = 20$) and for reward (11.20 vs. 6.62; $p < 0.05$; Mann-Whitney U-test, $N = 20$). Whereas this has been expected for punishment (see Proposition 5), the same is not true for reward (see Proposition 6).

Result 2. *When the institution is fixed exogenously and when punishment or reward are highly effective (|L| = 3), both institutions trigger significantly higher contribution levels than the standard linear VCM or the punishment or reward institution under |L| = 1.*

Checking profits and thus the important issue of the efficiency of different institutions under $|L| = 3$, panel (B) of Table 2 reveals that average profits are significantly higher under reward (29.37) than under both punishment (24.83) and the standard VCM (23.90). The underlying reason for this is that reward with high effectiveness is efficiency enhancing, whereas actual punishment has rather high social costs (of 4 ECU per punishment). The latter fact explains why profits with punishment under $|L| = 3$ are not higher than in the standard VCM even though average contributions are almost twice as high.

[18] All tests reported in this paper are two-sided, and the overall average contribution in each group is always treated as a single, statistically independent observation.

Turning to the frequency of using the punishment and reward options in the exogenous treatments, we can provide evidence that both are used more often in the high leverage condition $|L| = 3$ than in the low leverage condition, but only significantly so with reward. Rewards are assigned in 44% of all possible cases under $|L| = 3$ and only in 10% of cases under $|L| = 1$ ($p < 0.05$; Mann-Whitney U-test, $N = 20$). This is a clear indication that subjects exploit the potential efficiency gains under $|L| = 3$ to a large extent. The leverage has, however, no significant effect on the frequency of punishment, which is used in 20% of all cases under $|L| = 3$ and in 15% under $|L| = 1$.

Result 3. *Regarding the actual frequency of using punishment or reward, subjects are much more responsive to a higher leverage in the reward institution than in the punishment institution. Under $|L| = 3$ rewarding is much more frequent than under $|L| = 1$.*

Figure 2 and Figure 3 about here

Figures 2 and 3 provide a more detailed picture of the patterns of reward and punishment, depending upon the leverage. They show the relative frequency with which a subject *j* is rewarded or punished by subject *i*, contingent on the absolute difference between subject *i*'s and subject *j*'s contribution.[19] Concerning the frequency of punishment, we can discern from the left-most bar in Figure 2, for instance, that in 46% of the cases in which member *j* contributes 14 to 20 tokens less than member *i* we observe punishment of member *j* by member *i*. In general, there is a decrease in the relative frequency of punishment when moving from left to right in Figure 2. This means that members *j* are more frequently punished the less they contribute in comparison to member *i*. Like Fehr and Gächter (2000), we also observe a certain, but relatively small number of low contributors that punish high contributors.

The picture is different for rewards (in Figure 3), where one it is generally more likely to be rewarded the higher one's contribution is compared to the contribution of the

[19] Some other papers use the average group contribution as a reference point to study the patterns of punishment and reward instead of the pairwise comparison of contributions. While results are almost the same, the reason we opted for the second possibility is that we use the FS-model to derive theoretical prediction, which also builds upon individual comparisons, and we wanted to remain consistent with regard to that choice throughout the paper.

rewarding member. Under the high leverage, though, it seems that the subjects with high contributions reward each other, which leads to a peak in the central deviation range of [-2, 2] between member *j*'s and member *i*'s contributions. Finally, note that under the high leverage the probability of getting rewarded is generally very high if one contributes at least as much as the potentially rewarding subjects.

Result 4. *The relative frequency of punishment increases the larger the negative deviation from one's own contribution. Reward is most often granted for subjects that contribute at least as much or more than the rewarding subject.*

5.2 Behavior with endogenously determined institutions

5.2.1 Participation in the vote and institutional choice

Contrary to Prediction 2 and Proposition 4, which predicted a participation rate of zero, 44% of our subjects participate in the voting procedure in the low-leverage condition (i.e. 71 out of 160 subjects). In the high-leverage condition, 60% of subjects take part in the costly vote (i.e. 48 out of 80). As expected (see Proposition 7), voter turnout is significantly higher in the high-leverage condition ($\chi^2 = 5.21$, $p < 0.05$).

Result 5. *Voter turnout is significantly higher when reward and punishment are more effective (|L| = 3) than when they are less effective (|L| = 1).*

In Table 3, we report the number of voting rounds until a unanimous agreement was reached, contingent upon the number of group members that participated in the vote. Of course, all groups with only one voter finish the procedure after the first round because the single voter can implement his most desired institution. Across both leverage levels, 46 out of 57 groups reach an agreement within the first four voting rounds.[20] In the low-leverage condition, the maximum number of voting rounds is 26, whereas in the high-leverage condition three groups need more than 50 voting rounds to reach an agreement, with a maximum of 103 voting rounds in one group (which finally settled for the standard VCM).

[20] There were three groups (all with $|L| = 1$) in which none of the 4 members voted. Coincidentally, the random determination of the institution yielded one group each for the standard VCM, the punishment and the reward institution.

Table 3 about here

From Table 1, one can infer the chosen institutions in the two different leverage conditions. In the low-leverage condition ($|L| = 1$), 25 out of 40 groups choose the standard VCM, five groups opt for the VCM with punishment and ten groups agree on the VCM with reward. The distribution of chosen institutions is significantly different from a random one and also markedly different from the distribution in the high-leverage condition ($\chi^2 = 19.61$, $p < 0.01$).[21] With the high leverage, 17 out of 20 groups choose the VCM with reward and the remaining three groups the standard VCM. No group ever opts for the VCM with punishment in the high-leverage condition, contrary to our Proposition 7. The latter result obviously implies that subjects aim for the option to reward each other mutually when the reward option increases overall efficiency. Note that the equilibrium in which each group member reward every other group member is actually quite frequent with 21% of all cases (36 out of 170). Not surprisingly, there is not a single such case of mutual rewarding under the endogenously chosen reward with $|L| = 1$.

Result 6. *The distribution of institutional choices depends significantly on the leverage of reward and punishment. Given a high leverage, 85% of groups endogenously choose the VCM with reward, whereas with a low leverage the majority of groups (63%) agree on the standard VCM.*

In order to check the robustness of our results regarding the endogenous institutional choice and subjects' satisfaction with their chosen roles, we asked them at the end of the experiment whether they would vote again and, if so, for which institution they would vote if the experiment were to be continued. The data from this questionnaire reveal that 72% of voters (86 out of 119) would vote again for exactly the same institution. Only 13% of voters would abstain from voting, and 15% would switch to a different institution.[22]

---

[21] Note that there is no significant relationship between the number of voting rounds before reaching an agreement and the chosen institution.

[22] Half of those subjects who would abstain from voting had voted for the standard VCM in the beginning. The rest of those indicating to abstain were almost equally split between voters who had in the beginning preferred punishment or reward. Regarding voters who would switch to another institution, there is no clear pattern. Seven out of ten subjects who had voted for the standard VCM in the beginning would

The answers of non-voters are markedly different, because 61% of non-voters indicated that they would actively participate in the vote if the experiment were repeated. Most of them would have voted for punishment under $|L| = 1$ (27 out of 47) and for reward under $|L| = 3$ (22 out of 27). This clearly indicates that our results on the endogenous institutional choice would have been very similar and our conclusions corroborated if it had taken place in a later stage after several periods of experience with the VCM.

5.2.2 Contributions in the endogenous treatment

Figure 4 shows the time trend of contributions over the ten periods in the different treatments with endogenous institutional choice, and panel (A) of Table 4 reports the overall averages. In the low-leverage condition ($|L| = 1$), contributions are lowest in the standard VCM (5.48), intermediate in the VCM with reward (8.53) and clearly highest in the VCM with punishment (16.13). Each pairwise comparison yields significant differences ($p < 0.05$ in each case; Mann-Whitney U-tests). This means that with an endogenous institutional choice, the possibility of rewarding and punishing increases the level of cooperation significantly in comparison to the standard VCM, even though the leverage of reward and punishment is low. Punishment works best here, as it is associated with average contributions that are almost twice as high as with reward.

In the high-leverage condition, no group agrees on the punishment option. Groups with the reward institution have significantly higher contribution levels (14.59) than groups in the standard VCM (9.17) ($p < 0.05$; Mann-Whitney U-test, $N = 20$) or groups with reward under $|L| = 1$ (8.53; $p < 0.05$, $N = 27$).

Result 7. *When the institution is determined endogenously, we find for both leverages that reward leads to higher contributions than the standard VCM. The same holds true for punishment with the low leverage. Since there is not a single group that chose the punishment option under the high leverage, we cannot assess its effect on contributions there.*

---

switch to reward, the other three to punishment. Five out of seven subjects who had preferred reward initially would vote for punishment, the two other for control. Only one subject who had voted for punishment would vote for a different institution, namely reward. Detailed figures are available upon request.

Table 4 and Figure 4 about here

Average profits per group member and period are shown in panel (B) of Table 4. They are always significantly lower in the standard VCM than with punishment or reward. Under the low leverage, average profits are about 15% higher with punishment than with reward ($p < 0.05$; Mann-Whitney U-test; $N = 15$).

Figures 5 and 6 show the relative frequency of punishment and reward, depending on the difference in contributions of group members *i* and *j*. Since the patterns are very similar to the ones observed in Figures 2 and 3, we refer to the discussion of these figures in Section 5.1. Overall, the reward option is again more often used under $|L| = 3$ than under $|L| = 1$ (55% vs. 17%; $p < 0.05$; Mann-Whitney U-test, $N = 27$). In the punishment treatment under the low leverage, subjects use the punishment option only in 10% of all cases. Still, they can sustain the highest contribution levels.

Figures 5 and 6 about here

5.2.3 How voting relates to contributions

We have already shown that 44% (under $|L| =1$), respectively 60% (under $|L| =3$), of subjects participated in the vote on the institution. In this subsection, we analyze whether the decision to participate in the vote is systematically related to a subject's contributions in the public goods game. One intuitive conjecture would be that subjects who participate in the vote are those who care relatively more for the common interest of the group because they try to implement an institution they consider to be the best. If so, one might expect higher contributions from voters than from non-voters. Table 5 reports the results of a censored tobit regression where we regressed a member's contribution on her decision to participate in the vote (1 = voter), on the selected institution (separate dummies for punishment and reward), on the leverage (1 if the leverage is high), and on the number of voters in a group.

Table 5 about here

Column [1] of Table 5 presents the marginal effects of the independent variables and considers only the contributions in the very first period as the dependent variable. It shows that the dummy for voters is significantly positive, despite controlling for other institutional variables within a group. Voters contribute about 4 ECU more (which corresponds to 20% more in terms of their endowment) than non-voters. Hence, participation in the vote is a significant indicator of a subject's initial level of cooperation.

Interestingly enough, this voter-effect disappears very quickly, as it is obviously overpowered by the evolution of decisions within a group. The dummy for voters already ceases to be significant in period 2 (see column [2] of Table 5). The dummy also remains insignificant if we take a subject's overall average contributions over all ten periods as the dependent variable (see column [3]). It is comforting to note, though, that the other variables of interest (punishment, reward and leverage) always remain significantly positive in columns [2] and [3].[23]

Table 6 about here

In Table 6 we report a probit regression of a member's decision to punish or reward on the following variables: member *i* being a voter, the number of voters in the group, member *i*'s contribution to the public good, and the difference between member *i*'s and member *j*'s contributions in a given period. We consider all 10 periods in a panel, controlling for the dependencies within groups. We find that voters are much more likely to reward other group members than non-voters, both under $|L| = 1$ and $|L| = 3$. Voters do not have a higher probability than non-voters to punish other group members, though. The average contribution within a group has a significantly negative impact on the likelihood of punishment, and a significantly positive influence on the likelihood of reward. Furthermore, the likelihood of member *i* to punish member *j* is significantly decreasing with the difference between member *j*'s contribution and member *i*'s contribution, and the likelihood of rewarding is increasing with this difference.

Result 8. *Subjects who participate in the vote on the institution contribute much more to the public good than non-voters in the beginning of the experiment. From period 2 on this*

[23] A Poission regression model that accounts for the integer values of the dependent variable yields basically the same results as those presented in Table 5.

premium disappears due to interaction effects. Voters reward others more often than non-voters, but they do not punish more often than non-voters.

5.3 The effect of an endogenous institutional choice on contributions

A comparison of the corresponding cells in Table 2 and Table 4 shows immediately that average contributions are always higher in the endogenous treatments than in the corresponding exogenous treatments in case a reward or punishment option exists. The difference is most remarkable and significant for the VCM with punishment under $|L| = 1$, where contributions are on average 16.13 in the endogenous case and 8.79 in the exogenous case ($p < 0.01$; Mann-Whitney U-test; $N = 15$). In the VCM with reward, the differences are weakly significant (6.62 vs. 8.53 under $|L| = 1$; $p = 0.10$, $N = 20$; and 11.20 vs. 14.59 under $|L| = 3$; $p = 0.09$, $N = 27$), lending support to the conclusion that an endogenous institutional choice is a device that increases the level of cooperation in comparison to exogenously imposed institutions. This is summarized in our final result.

Result 9. *Contributions are higher when reward and punishment are endogenously chosen than when these institutions are implemented by an external authority.*

Result 9 seems all the more remarkable as it is independent of the number of voters in a group. In fact, there is no significant correlation between the average contribution and the number of voters in a group, as the partial correlation coefficient – controlling for leverage and institution – is only 0.06 ($p > 0.3$). There is also no significant relationship between the number of voting rounds before implementing an institution and the average contribution in a group. Furthermore, the relative frequencies of reward, or punishment, are not significantly different between the exogenous and the endogenous treatments, given the low leverage $|L| = 1$. Only for the high leverage do we find a higher relative frequency of reward in the endogenous treatment than in the exogenous one (55% vs. 44% ($p < 0.01$; Mann-Whitney U-test; $N = 27$). In sum, we regard the evidence as an indication that it is the endogenous choice of the reward or punishment option *per se* that positively affects the contribution levels.

6 Conclusion

Earlier studies have been able to demonstrate that an option to punish defectors and reward cooperators in public goods games has a positive influence on the level of cooperation (see, for example, Fehr and Gächter, 2000, 2002; Sefton et al., 2002). Since the reward or punishment options have typically been implemented exogenously by the experimenters, it has remained an open question as to whether subjects would actually choose such institutions if given the choice and whether the opportunity to choose the desired institution endogenously would have any additional impact on the level of cooperation.

In this paper, we have shown that there is a positive effect of an endogenous choice of a reward or punishment mechanism on the level of cooperation within groups. We call this effect a democratic participation rights-premium to cooperation. This premium adds to the cooperation-increasing effect of reward or punishment when imposed exogenously. We consider the demonstration of the existence of this premium the first important finding of our paper.

One possible explanation for the higher levels of cooperation with endogenous institutional choice might be due to voters *intentionally* choosing an institution. This might be perceived – by non-voters as well as other voters – as a signal that these voters actually plan to use the instruments at hand. Consequently, in the case of the VCM with punishment, the collective decision might pose a more credible threat to free-riders than the exogenous implementation. Furthermore, the threat is already salient in the first period to a greater extent, whereas with exogenous institutions it might take a few periods to deter free-riding by building up a punishment reputation. Hence, selfish subjects might be more reluctant to free-ride in the endogenous treatment. In the case of the VCM with reward, the signal from the vote might induce subjects to expect rewards in case they contribute relatively high amounts. In fact, the reward option is used significantly more often in the endogenous treatment than in the exogenous treatment when the leverage is high.

Therefore, from a behavioral perspective the endogenous choice of an institution seems to contain more than just the determination of a given set of rules. It also implies the signal of actively supporting – and probably applying – an available instrument. Given that the endogenous choice of the standard VCM does not provide group members with any instrument to express their (dis)satisfaction with the other members' contributions, it is

also not surprising that contributions in the standard VCM do not differ between the exogenous and the endogenous case.

One straightforward implication of our findings concerning the positive effects of an endogenous choice is that implementing an endogenous institutional choice may provide a welfare-enhancing substitute for increasing the leverage of punishment through an exogenous authority. Our results in the exogenous treatments as well as recent studies by Anderson and Putterman (2006), Carpenter (2006), Casari and Luini (2005), Egas and Riedl (2005), Fehr and Gächter (2003) or Nikiforakis and Normann (2005), for instance, have shown that contributions to a public good typically increase when the effectiveness of punishment is increased *exogenously*. However, a high leverage of punishment usually implies high social costs because it decreases overall efficiency. In our experiment we are able to provide evidence that instead of an increase in the effectiveness of punishment one might resort to an endogenous choice of the rules. Note that we find contributions in the *endogenous* treatment with $|L| = 1$ to be even significantly higher than contributions in the *exogenous* treatment with $|L| = 3$ ($p < 0.05$; Mann-Whitney U-test; $N = 15$).

Interestingly enough, cooperation levels in the endogenous treatments do not depend on the number of voters in a group (i.e. the voter turnout). That means that the higher *aggregate* cooperation under reward or punishment is induced through the mere opportunity of an endogenous choice, but does not depend on whether people actually participate in the vote. On the individual level, however, we have found that the voluntary (but costly) participation in the vote is a good indicator of a subject's level of cooperation at the beginning of the experiment. Voters have been estimated to contribute in the first period about 4 ECU (i.e. 20% of their endowment) more to the public good than non-voters. This is a noteworthy effect because it implies that the decision to participate in a vote can be linked to economic behavior. Voters, however, do not contribute significantly more than non-voters from the second period on, which is a possible indication that voters are conditional cooperators (Fischbacher et al., 2001) who react to the contributions of non-voters. Consequently, the behavior of voters and non-voters converges rather quickly after both types of subjects have interacted with each other. The latter result implies that the prevailing institutions overpower the vote-participation effect on the aggregate level.

We turn now to our results regarding the endogenous institutional choice. A second important finding of our paper is that the choice of institutions seems to depend heavily on their effectiveness. When reward or punishment have a high leverage, groups choose

almost exclusively the reward option. A straightforward explanation for this choice behavior might be the intention to exploit the possible efficiency gains through mutual rewards when reward is very effective. Another explanation for groups shying away from the punishment institution might be the fear of high efficiency costs if punishment is actually applied under $|L| = 3$.

When the leverage is low, however, a clear majority of groups prefer the standard VCM over the reward or punishment mechanism. The main reason for preferring the standard VCM might be the expectation of subjects that reward or punishment with $|L| = 1$ may have little or no effect at all, as is actually true when reward and punishment with the low leverage are implemented *exogenously*. If the institution is chosen endogenously, though, reward or punishment increase the contribution levels even with a low leverage. In fact, total profits have been significantly higher both with punishment and with reward than with the standard VCM (see Panel (B) of Table 4). Hence, from an ex-post perspective it would have paid off to vote for punishment or reward even with the low leverage.

Our results on the relative attractiveness of punishment versus reward under endogenous institutional choice have not been reported in the literature before. Neither Botelho et al. (2005) and Ertan et al. (2005) nor Gürerk et al. (2005) have been able to address this issue because Botelho et al. (2005) and Ertan et al. (2005) only considered punishment and Gürerk et al. (2005) did not allow for a choice between reward *and* punishment. Our results on the relative *un*attractiveness of punishment are clearly more in line with the findings of Botelho et al. (2005) than those of Gürerk et al. (2005), and they seem to be compatible with Ertan et al.'s (2005) finding that groups never allow punishment of high contributors. Botelho et al. (2005) argue that subjects opted for the standard VCM because subjects had experienced it as more efficient than the punishment institution. In our experiment, groups favored the reward institution over punishment possibly because of the higher efficiency of reward. Both findings should be considered complementary, indicating that subjects do not find the implementation of a punishment option attractive *if* they have an alternative. If they do not have any alternative like in the exogenous case, the punishment option is frequently used and has a positive effect on contributions (Fehr and Gächter, 2000, 2002), as long as the effectiveness of punishment is sufficiently high (Nikiforakis and Normann, 2005). Nevertheless, it is usually associated with ambiguous effects on overall efficiency, and as we know from many other

experiments, efficiency considerations play an important role in the decision-making of subjects (Charness and Rabin, 2002; Engelmann and Strobel, 2004). Therefore, ex post and from a behavioral perspective, the relative attractiveness of the reward mechanism seems fairly intuitive.

Regarding the theoretical predictions, it is not surprising that a model with purely selfish and rational agents is not consistent with our experimental results. The inequity aversion model of Fehr and Schmidt (1999), which we used to derive alternative predictions, fits our data much better, in particular with respect to the influence of the punishment option under the high leverage. Yet, it also fails to predict the observed behavior under the reward institution and the effects of endogenous institutional choice. One straightforward explanation for the latter failure is that the FS-model abstracts from intentions and from the emotional and physiological consequences of the act of punishing or rewarding (de Quervain et al., 2004). In fact, a large fraction of subjects seem to use the punishment and reward options either to reciprocate kind or unkind decisions of others or to express their (dis)satisfaction with the contributions of others, independently from the effects on the final distribution of profits. Such behavior can be called expressive punishment or expressive reward, and it can at best partly be captured by outcome-based models of social preferences. An analogous argument applies to the failure of explaining the democratic participation rights-premium. Going to the ballots and voting for a certain institution may reveal some information about a voter's intentions (on how) to use a certain institution. This information may influence the behavior of other group members, both voters and non-voters alike. Therefore, we consider it a worthwhile future research project to examine whether intention-based models (for example in the spirit of Rabin, 1993, Dufwenberg and Kirchsteiger, 2004, or Falk and Fischbacher, 2005) are better suited to explain both the influence of reward and the positive effects of endogenous institutional choice on behavior in social dilemma situations.

References

Anderson, C. M., Putterman, L. (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., Harbaugh, W., Vesterlund, L. (2003), The carrot or the stick: Rewards, punishments and cooperation. *American Economic Review* 93: 893-902.

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

Botelho, A., Harrison, G. W., Costa Pinto, L. M., Rutström, E. E. (2005), Social norms and social choice. University of Central Florida, Working Paper.

Brown, M., Falk, A., Fehr, E. (2004), Relational contracts and the nature of market interactions. *Econometrica* 72: 747-780.

Carpenter, J. (2006), The demand for punishment. *Journal of Economic Behavior and Organization*, forthcoming.

Casari, M. (2005), On the design of peer punishment experiments. *Experimental Economics* 8: 107-115.

Casari, M., Luini, L. (2005), Group cooperation under alternative peer punishment technologies: An experiment. University of Siena, Working Paper 2/2005.

Casari, M., Plott, C. R. (2001), Decentralized management of common property resources: experiments with a centuries-old institution. *Journal of Economic Behavior and Organization* 51: 217-247.

Charness, G., Rabin, M. (2002), Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117: 817-869.

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

Denant-Boemont, L., Masclet, D., Noussair, C. (2005), Anonymity in punishment, revenge and cooperation: A public good experiment. Emory University, Working Paper.

de Quervain, D., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., Fehr, E. (2004), The neural basis of altruistic punishment. *Science* 305 (27 August 2004): 1254-1258.

Dufwenberg, M., Kirchsteiger, G. (2004), A theory of sequential reciprocity. *Games and Economic Behavior* 47: 268-298.

Egas, M., Riedl, A. (2005), Cooperation and punishment in the Dutch: Evidence from a large internet experiment. University of Amsterdam, Working Paper.

Engelmann, D., Strobel, M. (2004), Inequity aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review* 94: 857-869.

Eriksson, T., Villeval, M.-C. (2004), Other-regarding preferences and performance pay. An experiment on incentives and sorting. Bonn, IZA Discussion Paper No. 1191.

Eriksson, T., Teyssier, S., Villeval, M.-C. (2005). Does self-selection improve the efficiency of tournaments? GATE, Working Paper.

Ertan, A., Page, T., Putterman, L. (2005), Can endogenously chosen institutions mitigate the free-rider problem and reduce perverse punishment? Brown University, Working Paper.

Falk, A., Fehr, E., Fischbacher, U. (2005), Driving forces behind informal sanctions. *Econometrica* 73: 2017-2030.

Falk, A., Fischbacher, U. (2005), A theory of reciprocity. *Games and Economic Behavior*, forthcoming.

Fehr, E., Fischbacher, U. (2004), Third party punishment and social norms. *Evolution and Human Behavior* 25: 63-87.

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

Fehr, E., Gächter, S. (2002), Altruistic punishment in humans. *Nature* 415 (10 January 2002): 137-140.

Fehr, E., Gächter, S. (2003), Self-governing solutions to collective action problems: Rewards versus punishments. Preliminary draft.

Fehr, E., Klein, A., Schmidt, K. (2001), Fairness, incentives and contractual incompleteness. Institute for Empirical Research in Economics, Working Paper No. 72, University of Zurich.

Fehr, E., Schmidt, K. (1999), A theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114: 817-868.

Fehr, E., Schmidt, K. (2004), Fairness and incentives in a multi-task principal-agent model. *Scandinavian Journal of Economics* 106: 453-474.

Feld, L., Tyran, J.-R. (2002), Tax evasion and voting: An experimental analysis. *Kyklos* 55: 197-221.

Fischbacher, U. (1999), z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual. Institute for Empirical Research in Economics, Working Paper No. 21, University of Zurich.

Fischbacher, U., Gächter, S., Fehr, E. (2001), Are people conditionally cooperative? Evidence from a public goods experiment. *Economic Letters* 71: 397-404.

Frey, B. S. (1994), Direct democracy: Politico-economic lessons from Swiss experience. *American Economic Review, Papers and Proceedings* 84: 338-342.

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

Gürerk, Ö., Irlenbusch, B., Rockenbach, B. (2005), On the evolvement of institution choice in social dilemmas. University of Erfurt, Working Paper.

Kirchsteiger, G., Niederle, M., Potters, J. (2005), Endogenizing market institutions: An experimental approach. *European Economic Review* 49: 1827-1853.

Kroll, S., Cherry, T. L., Shogren, J. F. (2005), Voting, punishment, and public goods. California State University Sacramento, Working Paper.

Lazear, E. P., Malmendier, U., Weber, R. A. (2005), Sorting in experiments. Stanford University and CMU, Working Paper.

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

Masclet, D., Villeval, M.-C. (2004), Is peer pressure in teams motivated by inequality aversion. University of Lyon, Working Paper.

Nikiforakis, N. (2005), Punishment and counter-punishment in public-good games: Can we really govern ourselves? Royal Holloway, University of London, Working Paper.

Nikiforakis, N., Normann, H.-T. (2005), A comparative statics analysis of punishment in public-good experiments. Royal Holloway, University of London, Working Paper.

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

Pommerehne, W. W., Weck-Hannemann, H. (1996), Tax rates, tax administration and income tax evasion in Switzerland. *Public Choice* 88: 161-170.

Rabin, M. (1993), Incorporating fairness into game theory and economics. *American Economic Review* 83: 1281-1302.

Rege, M., Telle, K. (2004), The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics* 88: 1625-1644.

Sefton, M., Shupp, R., Walker, J. (2002), The effect of reward and sanctions in the provision of public goods. University of Nottingham, Working Paper.

Yamagishi, T. (1986), The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology* 51: 110-116.

Tables and figures

Table 1: Treatment characteristics and number of independent observations

Determination of the institution	Leverage	Institution: Standard VCM	Institution: Punishment	Institution: Reward
Exogenous	$\|L\|=1$	$N = 10$[1)]	$N = 10$	$N = 10$
	$\|L\|=3$		$N = 10$	$N = 10$
Endogenous	$\|L\|=1$	$N = 25$	$N = 5$	$N = 10$
	$\|L\|=3$	$N = 3$	–	$N = 17$

[1)] There is no leverage in the standard VCM.

Table 2: Exogenous treatments

	Leverage	Standard VCM	Punishment	Reward
Panel (A) - Contributions	$\|L\|=1$	6.50[#,*] ($N = 10$)	8.79 ($N = 10$)	6.62 ($N = 10$)
	$\|L\|=3$		12.11[#] ($N = 10$)	11.20* ($N = 10$)
Panel (B) – Profits per member and period	$\|L\|=1$	23.90*	24.37	23.97
	$\|L\|=3$		24.83[§]	29.37*[,§]

[#] significant difference ($p < 0.05$) between standard VCM and punishment

* significant difference ($p < 0.05$) between standard VCM and reward

[§] significant difference ($p < 0.05$) between punishment and reward

Table 3: Number of voting rounds

Panel (A) – Low leverage

Number of voters in a group (*N* = total number of groups)	Reaching an unanimous vote in voting round …								
	1	2	4	7	9	10	11	13	26
1 (*N* = 15)	15								
2 (*N* = 13)	9	1		1		1		1	
3 (*N* = 6)	3	1					1		1
4 (*N* = 3)		1	1		1				

Panel (B) – High leverage

Number of voters in a group (*N* = total number of groups)	Reaching an unanimous vote in voting round …							
	1	2	3	11	13	51	61	103
1 (*N* = 5)	5							
2 (*N* = 6)	5		1					
3 (*N* = 5)	2	1			1		1	
4 (*N* = 4)			1	1		1		1

Table 4: Endogenous treatments

	Leverage	Standard VCM	Punishment	Reward
Panel (A) - Contributions	|*L*|=1	5.48[#,*] (*N* = 25)	16.13[#,§] (*N* = 5)	8.53*[,§] (*N* = 10)
	|*L*|=3	9.17* (*N* = 3)	-	14.59* (*N* = 17)
Panel (B) – Profits per member and period	|*L*|=1	23.29[#,*]	29.25[#,§]	25.12*[,§]
	|*L*|=3	25.51*	-	32.05*

[#] significant difference ($p < 0.05$) between standard VCM and punishment

* significant difference ($p < 0.05$) between standard VCM and reward

[§] significant difference ($p < 0.05$) between Punishment and reward

Table 5: Vote participation and contributions in the public goods game (censored tobit)

	Dependent variable: Contributions in …		
Independent variables	first period	second period	overall (Periods 1-10)
Voter (= 1)	4.01*	2.67	0.95
Punishment institution (= 1)	5.98**	15.44**	11.91**
Reward institution (= 1)	5.35*	10.98**	4.15**
Leverage (= 1 if $\|L\| = 3$)	-2.09	0.35	5.79**
Number of voters in group	0.94	0.15	-0.12
Intercept	3.84*	1.30	4.73*

** significant at 1% level; * significant at 5% level

Table 6: Punishing or rewarding behavior (panel probit-regression)

	Dependent variable: Decision to…		
Independent variables	punish with |L| = 1	reward with |L| = 1	reward with |L| = 3
Voter (= 1)	-0.020	0.051*	0.134**
Number of voters in group	0.027*	0.060**	-0.021
Group contribution	-0.001**	0.001**	0.005**
Difference between other member's and own contribution	-0.011**	0.008**	0.009**
Observations	600	1200	2040

** significant at 1% level; * significant at 5% level

Figure 1. Average contributions in the exogenous treatments

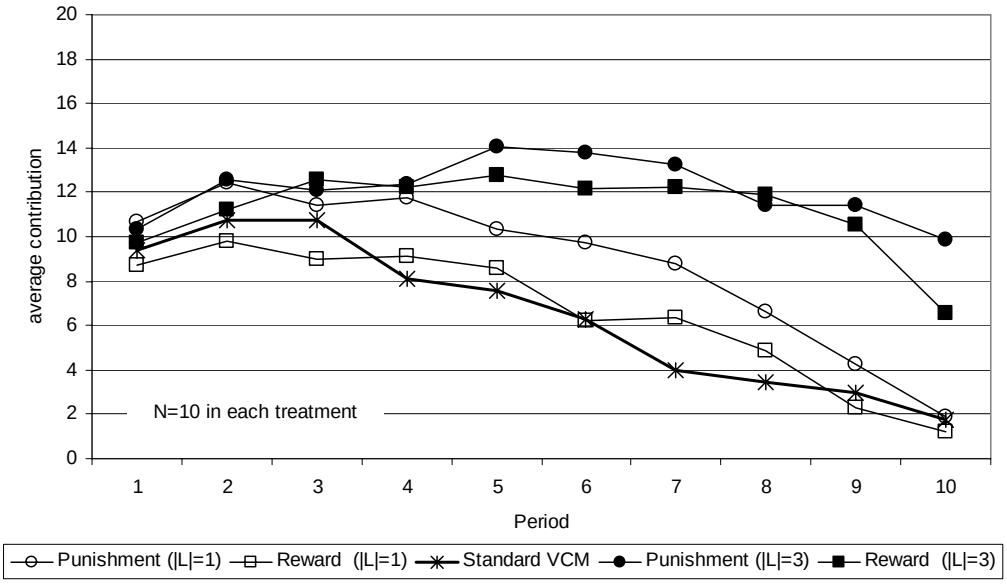
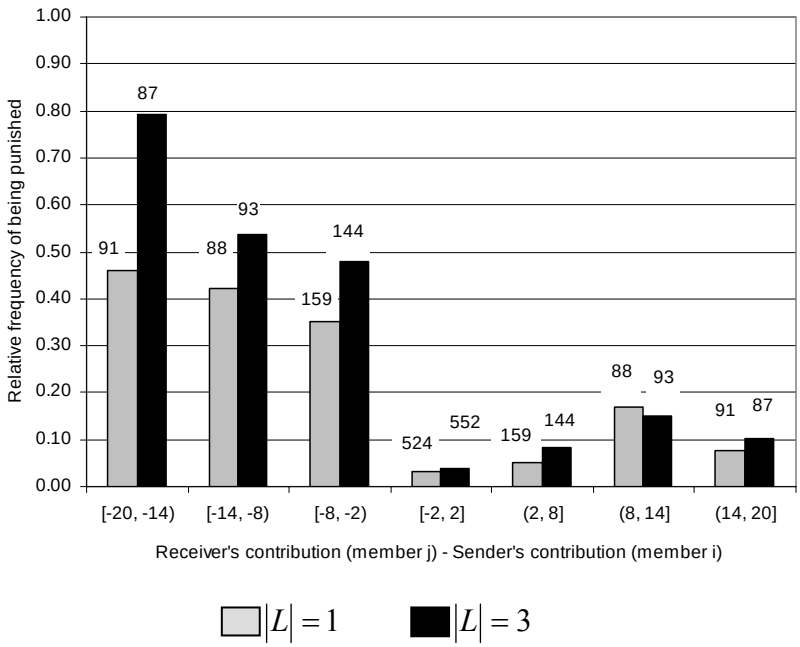
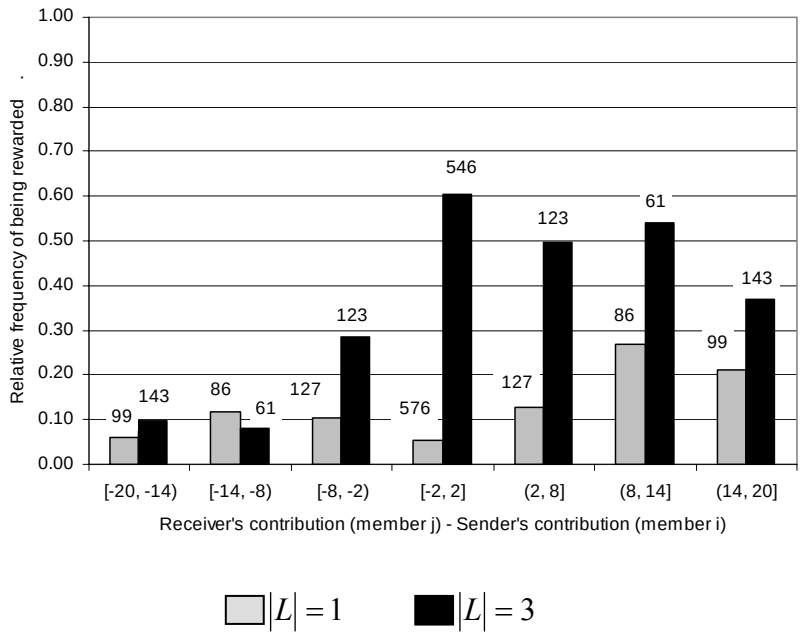


Figure 2. Punishment in the exogenous treatments



Numbers above bars indicate the number of observations.

Figure 3. Reward in the exogenous treatments



Numbers above bars indicate the number of observations.

Figure 4. Average contributions in the endogenous treatments

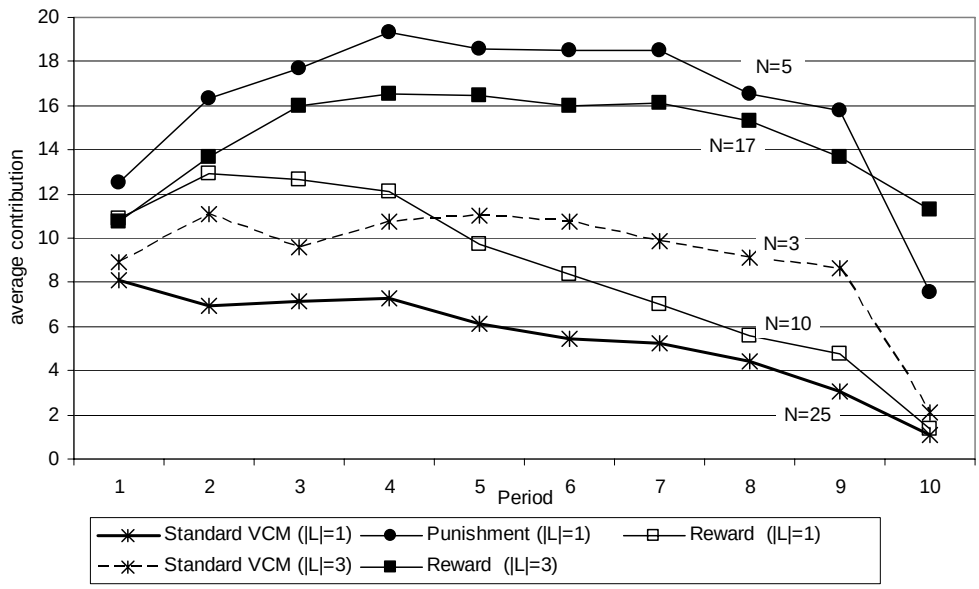
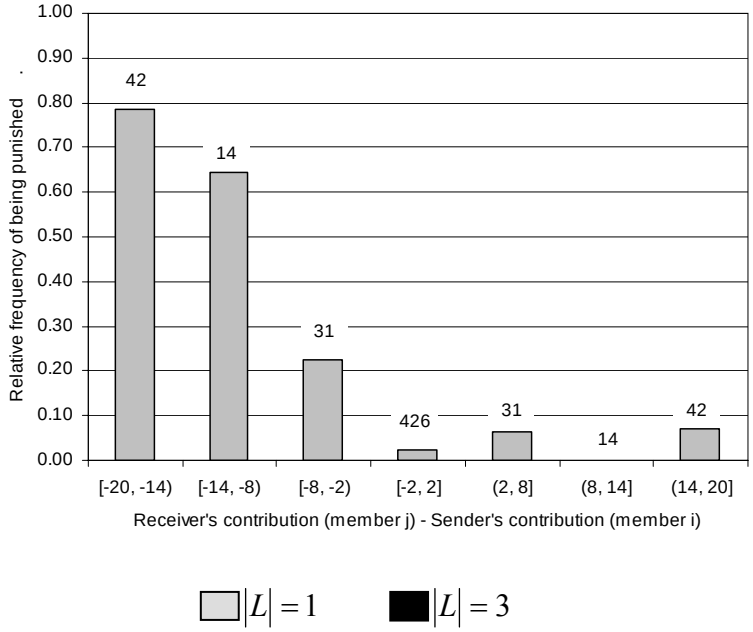


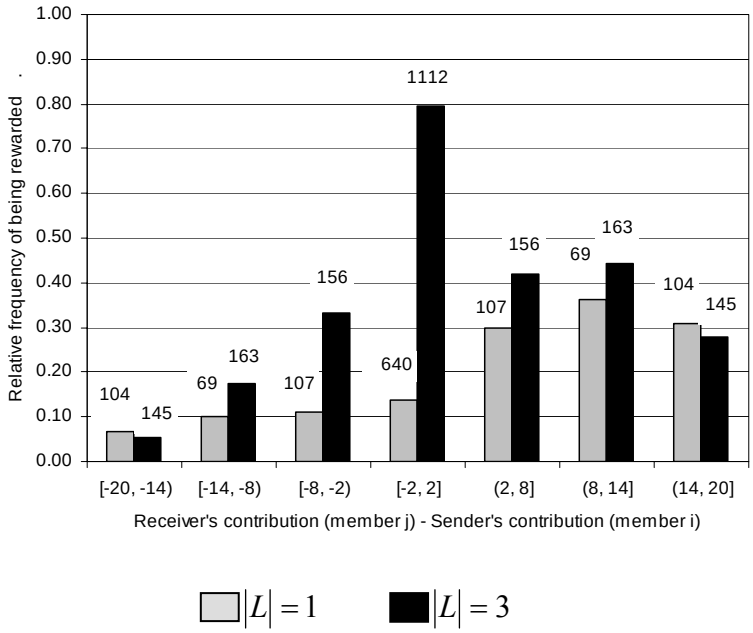
Figure 5. Punishment in the endogenous treatments



Numbers above bars indicate the number of observations.

Note: There are only data for $|L|$=1 because punishment was not chosen with $|L|$=3.

Figure 6. Reward in the endogenous treatments



Numbers above bars indicate the number of observations.

Appendix A: Instructions for the endogenous choice with $|L| = 3$

(Originally in German – the other instructions are available upon request**)**

Welcome to the experiment. Please refrain from talking to other participants from now on.

Groups of 4 persons and 10 periods

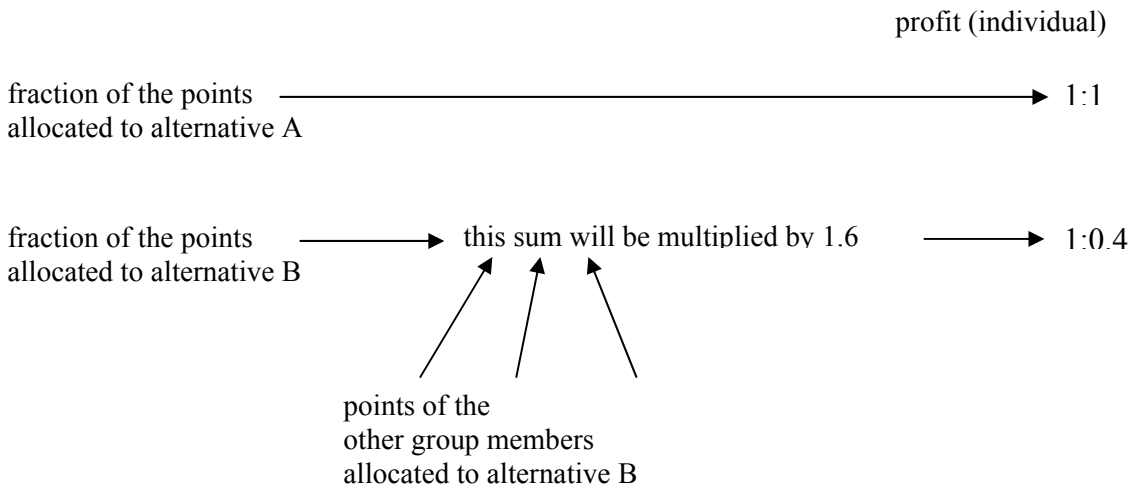
At the beginning of the experiment you are randomly assigned to a group of 4 subjects that will remain the same throughout the whole experiment. The whole experiment lasts 10 periods.

Basic decision

At the beginning of each period you will receive an endowment of 20 tokens. These tokens can be allocated into two alternatives, denoted A and B. The sum of the tokens allocated to A and B must equal 20. You cannot save any tokens or transfer them to a later period.

Profits from both alternatives

You will receive the tokens allocated to alternative A back one-to-one as one part of your profit.
The tokens allocated to alternative B are summed up with the contributions of the other group members to alternative B. The sum of all tokens allocated to alternative B in your group is multiplied by 1.6 and will be equally distributed– independent of one's particular contribution – among all group members. This means that you – and every other group member – will receive 0.4 tokens in return (=1.6/4) for every token allocated to alternative B.
The following figure summarizes the profits from both alternatives. The sum of profits from both alternatives gives your total earnings in this period.



Three possible institutions

Institution number one

Under this institution you have to make your basic decision only. Your profit per period is as described above, depending on how much you have allocated to alternative A and how much the group altogether has allocated to alternative B. After the basic decision has been entered by all subjects, you will be informed on the next screen about the decisions of all other group members.

Institution number two

This is the same as institution number one, with the following additional feature: After having observed the other members' contributions you have the option to assign points to other members. You can assign points to as many other members as you wish, but you can only assign one point per member. Assigning a point has the following consequences: The person assigning a point bears the cost of one token (this cost will be subtracted from the period payoff). The person receiving a point bears the cost of three tokens (also to be subtracted from her period payoff).

Institution number three

This is also the same as institution number one, with the following additional feature: After having observed the other members' contributions you have the option to assign points to other members. You can assign points to as many other members as you wish, but you can only assign one point per member. Assigning a point has the following consequences: The person assigning a point bears the cost of one token (this cost will be subtracted from the period payoff). The person receiving a point receives three tokens (to be added to her period payoff).

Which institution is valid for your group?

Before the experiment starts, you have the option to participate in a vote on the institution that will be valid in your group for all 10 periods. You are not required to participate in the vote, but even in that case the voting outcome in your group will be binding for you as well. Participation in the vote has a one-time cost of 10 tokens (which will be subtracted from your first period's earnings).

If you participate in the vote you have to indicate for each of the three institutions introduced above whether you support this institution (click on "Yes") or not (click on "No"). You can support as many institutions as you want. Before the vote starts, you will see the number of voters in your group. When all group members that participate in the vote have made their decisions, you will see on a succeeding screen how many group members supported any of the three institutions.

1. If one and only one institution is unanimously supported by all voters in your group, then this institution will be applied for the 10 periods.

2. If two or three institutions are unanimously supported, then the institution will be chosen randomly from those institutions unanimously supported by all voters in your group.
3. If no institution has unanimous support, then the voting procedure will be repeated.

The voting procedure is repeated until the unanimous support of at least one institution is reached.

If no group member participates in the vote, the institution is chosen randomly with equal probabilities for each institution.

[In order to speed up the voting procedure, we ask you to pay close attention to the status bar where you will see either the message "waiting for others" or "please vote".]

Screen sequences

In the upper part of the screen you will see your group and subject number. On all screens you will also be able to see the prevalent institution in your group.

Independent of the institution used, you always have to make the basic allocation decision. Then, you will be shown the basic decisions of all group members on the next screen. Under institution one you will also see the profits. Under institutions two and three you can then enter the points if you wish to assign some. As the final screen in institutions two and three of each period, you will see again the basic decisions of all group members (linked to their ID) and the sum of points you have assigned and received. You are not able to see how many points other group members have assigned or received or who has assigned points to you. The profits are calculated as described above.

Exchange rate. At the end of the experiment all tokens you have earned are summed up and converted in real money at a rate of **10 tokens = 0.40 EURO**.

Appendix B: Predictions and propositions

B1. Theoretic predictions when agents are purely selfish

B1.1 Standard voluntary contribution mechanism (Prediction 1)

The payoff for member i in period t is given by $\pi_{i,t} = E - c_{i,t} + \gamma C_t$, where $C_t = \sum_{j=1}^{n} c_{j,t}$ denotes the sum of contributions within a group. Given that $0 < \gamma < 1 < n\gamma$ it follows that the marginal return for member i from investing into the public good is negative (since $\partial \pi_{i,t} / \partial c_{i,t} = -1 + \gamma < 0$). Under the assumptions of selfishness and common knowledge of rationality, the only subgame-perfect equilibrium is to contribute zero in each single period (i.e. $c_{i,t} = 0 \; \forall t$).

B1.2 VCM with punishment or reward (Prediction 1)

Since punishment (reward) is costly (both with $|L|$ =1 and $|L|$ = 3), no member will ever punish (reward) another member in the second stage of the final period. Hence, contributions in the first stage of the final period cannot be affected by the availability of punishment (reward). Unraveling the same logic back to the first period yields zero contributions as the only subgame-perfect equilibrium. Given that equilibrium contributions are identical under all available institutions, it follows that all members should abstain from voting in order to save the voting costs (**Prediction 2**).

B2. Taking social preferences into account

The subsequent analysis assumes social preferences of the Fehr-Schmidt type (see equation (3) in section 4.2), and common knowledge of rationality.

B2.1 Standard voluntary contribution mechanism (Proposition 1)

For the proof of the conditions mentioned in section 4.2.1 we ask the reader to refer to the appendix of the paper by Fehr and Schmidt (1999, pp. 860ff.).

B2.2 Low leverage (|*L*| = 1)

B2.2.1 Punishment (Proposition 2)

Let us denote the number of cooperators (who satisfy $\gamma + \beta_i \geq 1$) in a group by *n'*. For reasons of generality, let us define *k* as the costs of punishing or rewarding another subject. Note that *k* is normalized to unity (*k* = 1) in the experiment. Consider a strategy where all members contribute *c* and where all cooperators punish any member who contributes $c_{i,t} < c$. For this strategy to be an equilibrium we must show that (*i*) it does not pay for any member to free-ride and contribute less than *c* and that (*ii*) cooperators have an incentive to punish those who contribute $c_{i,t} < c$, i.e. that the punishment threat is credible.

(*i*) Free-riding on the other members' contributions *c* by choosing $c_{i,t} < c$ generates a monetary gain of $(c - c_{i,t})(1-\gamma)$ for the free-rider (relative to those contributing *c*). If the *n'* cooperators punish the free-rider, the latter suffers a monetary loss of *n'L*. The maximum gain from a deviation from *c* is given for $c_{i,t} = 0$. The resulting gain is smaller than the loss from being punished as long as the following condition is satisfied.

$$c \leq \bar{c} = \frac{n'L}{(1-\gamma)}. \tag{B1}$$

Hence, no member has an incentive to deviate from contributing *c* as long as the latter condition is fulfilled. What remains to be shown is whether the threat of punishing those members who would consider contributing $c_{i,t} < c$ is credible.

(*ii*) For the threat to be credible, a cooperator's utility from punishing must be larger than her utility from not punishing, under the assumption that the other $(n'-1)$ cooperators stick to their punishment strategy. Hence, we have to check the following inequality for member *i*.

$$-k - \frac{\alpha_i}{n-1}(n-n'-1)k - \frac{\alpha_i}{n-1}(\bar{c} - c + k - n'L) \geq -\frac{\alpha_i}{n-1}(\bar{c} - c - (n'-1)L) - \frac{\beta_i}{n-1}(n'-1)k \tag{B2}$$

The first term on the left hand side denotes the costs of punishing a deviating member. The second term indicates the disadvantageous inequality towards those members who contribute *c*, but who do not punish deviating members. The third term captures the remaining disadvantageous inequality towards the deviating member who gets punished by all *n'* cooperators. The first term on the right hand side denotes the disadvantageous inequality towards the deviating member if member *i* does not punish, but the other *n'* – 1

cooperators punish. The second term is due to the advantageous inequality of member *i* towards the *n'* – 1 cooperators who punish. Rearranging and simplifying yields the following condition to be satisfied in order to make punishment a credible threat.

$$\frac{L}{k} \geq (n-n') + \frac{1}{\alpha_i}\left[(n-1) - \beta_i(n'-1)\right] \tag{B3}$$

Note that the left-hand side of equation (B3) yields *L/k* = 1 in our experiment in the low-leverage treatment (|*L*| = 1). From that it follows that condition (B3) cannot be satisfied as long as $n' < n$. Hence, the threat of punishment is not credible under |*L*| = 1 if *n'* < *n*. As a consequence, the same predictions as in the standard VCM apply. Only if $n' = n$, i.e. if all group members are cooperators (with $\gamma + \beta_i \geq 1$), the threat of punishment is credible and any contribution level $c \leq \bar{c}$ can be enforced.

B2.2.2 Reward (Proposition 3)

Suppose all members contribute $c \in [0, E]$. Note that the *n* – *n'* members with $\gamma + \beta_i < 1$ will never reward other members. A cooperator, then, has to consider the following inequality in order to decide whether to reward (left-hand side) or not (right-hand side).

$$(n'-1)L - (n-1)k - \frac{\alpha_i}{n-1}\left[L + (n-1)k\right](n-n') > (n'-1)L - (n'-1)\frac{\beta_i}{n-1}[L-(n'-1)k] \tag{B4}$$

The first term on the left hand side captures the gains from being rewarded by the other *n'* – 1 cooperators. The second term shows the costs from rewarding all other members. The third term denotes the disadvantageous inequality towards the *n* – *n'* members with $\gamma + \beta_i < 1$ who do not reward others. The first term on the right hand side shows again the gains from being rewarded by the other cooperators. The second term indicates the advantageous inequality towards those *n'* – 1 cooperators. Rearranging terms yields the following condition that has to be satisfied in order to make cooperators reward the other members.

$$(n-1)k + \frac{1}{n-1}\left[L + (n-1)k\right]\left[\alpha_i(n-n') - \beta_i(n'-1)\right] < 0 \tag{B5}$$

This condition is never satisfied for *n'* < *n*. Therefore, there are no equilibria in which reward is part of the equilibrium strategy. Thus, for $n' < n$ we have the same set of equilibria as in the standard VCM.

For $n' = n$ we find that mutual rewarding can be part of an equilibrium strategy. To show this, we examine a situation where all members contribute *c* and reward each other. If member *i* deviates and does not reward the other $(n-1)$ members, she saves rewarding costs of $(n-1)k$. But at the same time member *i* suffers from advantageous inequality, expressed by $\frac{\beta_i}{n-1}[L+(n-1)k](n-1)$. Sticking to reward is thus optimal if

$$(n-1)k \leq \frac{\beta_i}{n-1}[L+(n-1)k](n-1) \tag{B6}$$

For |*L*| = 1 and *k* = 1 we must have $\beta_i \geq 0.75$ for condition (B6) to hold. If all group members satisfy $0.6 < \beta_i < 0.75$, any positive contribution $c_{i,t} = c \in [0, E]$ can still be sustained as an equilibrium, but mutual reward is not part of an equilibrium.

Given the positive costs of voting, Propositions 2 and 3 imply that subjects do not participate in the vote when |*L*| = 1 (**Proposition 4**).

B2.3 High leverage (|*L*| = 3)

B2.3.1 Punishment (Proposition 5)

Under the high-leverage |*L*| = 3 condition (B3) can be satisfied even if $n' < n$ (depending on a member's α_i and β_i). Hence, punishment can be a credible threat even if not all group members are cooperators. Therefore, it is possible to enforce $\bar{c}$. All other equilibria discussed for the case of punishment with |*L*| = 1 remain also valid for |*L*| = 3.

B2.3.2 Reward (Proposition 6)

From condition (B5) it can be seen that even under the high leverage treatment it is not possible to enforce any $c \in [0, E]$ through the use of rewards since it is better even for cooperators to abstain from rewarding as long as $n' < n$. Only if $n' = n$ mutual rewarding can be part of an equilibrium strategy. From equation (B6) and the assumption that $n' = n$ it follows that mutual rewarding is part of an equilibrium strategy already if $\beta_i \geq 0.6$ for all members (which is a less restrictive condition than under |*L*| = 1 where it was necessary to satisfy $\beta_i \geq 0.75$ for all members). Propositions 5 and 6 imply that subjects should vote for punishment with |*L*| = 3 (**Proposition 7**).