

# Public goods provision and sanctioning in privileged groups

Citation for published version (APA):

Reuben, E., & Riedl, A. M. (2007). *Public goods provision and sanctioning in privileged groups*. METEOR, Maastricht University School of Business and Economics. METEOR Research Memorandum No. 028 <https://doi.org/10.26481/umamet.2007028>

## Document status and date:

Published: 01/01/2007

## DOI:

[10.26481/umamet.2007028](https://doi.org/10.26481/umamet.2007028)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Ernesto Reuben, Arno Riedl

Public Goods Provision and Sanctioning in  
Privileged Groups

RM/07/028

JEL code: H41, D01, C92, Z13



Maastricht research school of **E**conomics  
of **T**Echnology and **O**Rganizations

Universiteit Maastricht  
Faculty of Economics and Business Administration  
P.O. Box 616  
NL - 6200 MD Maastricht

phone : ++31 43 388 3830  
fax : ++31 43 388 4873



# Public Goods Provision and Sanctioning in Privileged Groups\*

Ernesto Reuben<sup>†</sup> and Arno Riedl<sup>‡</sup>

July 5, 2007

## Abstract

In public good provision, privileged groups enjoy the advantage that some of its members find it optimal to supply a positive amount of the public good. However, their inherent asymmetric nature may make the enforcement of cooperative behavior through informal sanctioning harder to accomplish. In this paper we experimentally investigate public good provision in normal and privileged groups with and without decentralized punishment. We find that compared to normal groups, privileged groups are relatively ineffective in using costly sanctions to increase everyone's contributions. Punishment is less targeted towards strong free-riders and they exhibit a weaker increase in contributions after being punished. Thus, we show that privileged groups are not as privileged as they initially seem.

JEL Codes: H41, D01, C92, Z13

Keywords: privileged groups, public goods, punishment, cooperation, collective action,

---

\*We would like to thank Enrique Fatas, Dan Levin, Jan Potters, Louis Putterman, Frans van Winden, and particularly Simon Gächter for useful comments and suggestions. Financial support from the Dutch Science Foundation (NWO) through the "Evolution & Behavior" grant 051-12-012 and from the EU-Marie Curie RTN ENABLE (MRTM-CT-2003-505223) is gratefully acknowledged.

<sup>†</sup>Northwestern University, e-mail: ereuben@northwestern.edu

<sup>‡</sup>CESifo, IZA, and Maastricht University, e-mail: a.riedl@algec.unimaas.nl

# 1 Introduction

In public good provision the incentives to free-ride on the contributions of others generally lead to under-provision of the public good. In his seminal work on collective action, Olson (1965) provides a typology of groups that helps to explain why some groups nevertheless perform better than others in collective good provision. In particular, he identifies a type of group where at least one member “has an incentive to see that the collective good is provided, even if he has to bear the full burden of providing it himself” (Olson, 1965, p. 50). In such groups one or more individuals receive a marginal benefit from the public good that exceeds their marginal cost of providing it. Hence, in such groups at least some level of the public good is voluntarily produced. Correspondingly, Olson also calls them privileged groups.

The theoretical literature on collective action generated important insights about when groups can be regarded as privileged. It has identified important determinants that turns non-privileged groups into (fully) privileged ones and, hence, circumstances that mitigate under-provision of public goods (see e.g., Sandler, 1992). Perhaps, because under-provision of public goods seems less problematic in privileged groups, considerably less attention has been devoted to this issue. A main purpose of this paper is to study precisely this problem and show that privileged groups may not always be as privileged as they seem.

In our view, there are at least three reasons that call for further exploration of voluntary public goods provision in privileged groups. First, it is clear that many groups confronted with the problem of providing public goods are privileged in the sense of Olson (1965). Contributing to the fight against international terrorism is an obvious example. The United States and the United Kingdom, for instance, can be assumed to contribute because they are likely targets and, hence, their governments perceive the benefits of contributing as being larger than the costs. In contrast, some continental European countries may see themselves as unlikely victims, and hence, perceive the costs as being larger than the benefits. The same could be said of most cases of international security such as containment of ‘rouge’ regimes. Another example is the so called commons-based peer-production model where a large number of anonymous people with varying degrees of interest voluntarily contribute to a common project.<sup>1</sup>

Second, even though there are some voluntary contributions in privileged groups, under-provision will occur as long as not all individual incentives are aligned with social welfare maximization.<sup>2</sup> That such under-provision is not only a theoretical issue is vividly revealed

---

<sup>1</sup>The probably best known examples are the GNU/Linux operating system and the WWW encyclopedia Wikipedia. Benkler (2002) discusses a number of other examples.

<sup>2</sup>Under-provision is a particular problem for the mostly assumed summation and weakest-links technologies of

in national and international political discussions regarding contributions to fighting terrorism and global warming, for instance. As for ordinary, non-privileged, groups this asks for studying appropriate mechanisms overcoming the problem of sub-optimal collective good provision.

Third, experimental research has recently identified important informal determinants of voluntary contributions, like fairness oriented and reciprocal inclinations. These determinants, however, may work out differently in privileged groups. The asymmetric nature of such groups may cause conflicting perceptions of fairness and reciprocity norms (Knez and Camerer, 1995), which in turn may undermine the otherwise very successful enforcement mechanism of decentralized individual punishment (Ostrom et al., 1992; Fehr and Gächter, 2000, 2002).

This paper is the first to experimentally investigate voluntary contributions to a public good in privileged groups vis-à-vis normal, non-privileged, groups. In particular, we compare behavior of subjects with different incentives to contribute in situations with and without the option to costly punish free-riders. Our results indicate that, without punishment, there is substantial under-provision in privileged groups, albeit contributions are higher than in normal groups. With punishment, privileged groups lose their privileged status completely. Contributions to the public good are not higher than in non-privileged groups and there is still substantial under-provision. In addition, inequality in privileged groups increases significantly when punishment is possible. Hence, the otherwise powerful institution of decentralized punishment loses some of its bite in privileged groups. The identification of the underlying reasons for this observation is another aim of our paper.

We conducted a linear public goods experiment (Isaac et al., 1984) where in some treatments subjects are allowed to punish each other after having observed individual contributions to the public good (similar to Fehr and Gächter, 2000). Subjects participate in either a privileged or a normal group. In normal groups, material incentives are such that all subjects have a dominant strategy to free ride. In privileged groups, there is one subject who's material incentives are such that full contribution is a dominant strategy whereas other group members' dominant strategy is to free-ride completely. For convenience, we refer to subjects with the different material incentives to contribute as high- and low-benefit subjects.<sup>3</sup>

---

publicness. For public goods provided with a best-shot technology, privileged groups would, in theory, overcome to a large extent the under-provision problem (see Sandler, 1992).

<sup>3</sup>The reasons why individuals may value consumption of the public good differently are numerous. It could be simply a difference in monetary returns. For instance, an individual that owes a large plot of land would benefit much more from a regional irrigation system than an individual that owes only a small plot of land. It may also be that people perceive the importance of the public good differently. For instance, some neighbors might enjoy more than others the existence of a neighborhood swimming pool. Voluntary contributors in commons-based

Note, that under the standard assumption of (common knowledge of) narrow selfish preferences only high-benefit subjects are predicted to contribute (fully) to the public good. In addition, neither high- nor low-benefit subjects are predicted to punish. However, there is plenty of evidence that in normal groups subjects, first, have some (initial) inclination to contribute positive amounts and, second, that individual costly punishment is undertaken and effective in decreasing free-riding.<sup>4</sup> These behavioral regularities are often attributed to subjects following reciprocity-based and/or outcome-based fairness norms. From the outset, however, it is unclear if such norm guided behavior can also successfully increase contributions in privileged groups. First, since high-benefit subjects benefit disproportionately from contributing to the public good it is unclear if such behavior should be interpreted as kind towards others or as barely selfish. Since intentions have been shown to matter (Falk et al., 2000; McCabe et al., 2003; Charness and Levine, 2006) such ambiguity may make low-benefit subjects unwilling to reciprocate high contributions of high-benefit subjects, and high-benefit subjects unwilling to punish low contributions of low-benefit subjects. Second, since contributions of low-benefit subjects can increase income differences with high-benefit subjects, earnings-based notions of fairness support outcomes where low-benefit subjects do not contribute to the public good.<sup>5</sup>

Our results indeed show that differences in punishment as well as in the response to being punished can explain why privileged groups fail to outperform normal groups. The main reason is that the punishment institution produces a much smaller increase in contributions by low-benefit subjects in privileged groups. This in turn can be attributed to relatively unstructured punishment behavior of high-benefit subjects in combination with a weak response in contributions to being punished by low-benefit subjects.

Experimental studies have investigated the role of incentives to cooperate (i.e, the marginal per capita return to contribute) in public goods settings quite extensively. From these studies it is well known that changes in the marginal per capita return has significant effects on contribution organizations are another example. They likely receive (perceived or expected) benefits in excess of the costs from contributing to the project. The benefits may range from pure pleasure from contributing to the project to higher likelihoods of getting a contract in traditional firms with experience in such commons-based organizations.

<sup>4</sup>Evidence of voluntary contribution without punishment can be found in Ledyard (1995). Examples of studies on decentralized punishment include Ostrom et al. (1992), Fehr and Gächter (2000), Masclet et al. (2003), Egas and Riedl (2005), Gächter and Herrmann (2006), Bochet et al. (2006), and Carpenter (2006).

<sup>5</sup>The same holds for other types of distributional fairness notions such as a concern for the income of the least well off Rawls (1971). On the other hand, if subjects care for overall efficiency, low contributions by low-benefit subjects are clearly undesirable. For more discussion on distributional preferences see Charness and Rabin (2002) and Engelmann and Strobel (2004).

tion behavior (see Ledyard, 1995). A few studies have made cooperating the dominant strategy (Saijo and Nakamura, 1995; Palfrey and Prisbrey, 1996, 1997). The studies most closely related to this paper are Fisher et al. (1995) and Brandts and Schram (2001). In both studies, the effect of unequal incentives to contribute within a group is explored. However, none of these studies investigates privileged groups as we do.

The rest of the paper is organized as follows. In Section 2 we describe the design of the experiment and derive qualitative predictions for theoretical models assuming fairness preferences and reciprocity. In Section 3 we present the results concerning subjects' contribution and punishment behavior. Section 4 closes with a summary of our main results and concludes.

## 2 Experimental Design

The basic game implemented in the experiment is a linear public goods game. Each subject participates in only one of four experimental treatments. The treatments differ in two dimensions: first, the public good game is either followed by a decentralized punishment stage a la Fehr and Gächter (2000) (*punishment* condition) or there is no such punishment stage (*baseline* condition); second, subjects in a group differ with respect to the marginal benefit they receive from the public good (*privileged* group condition) or do not differ in this respect (*normal* group condition). The game is played by the same group of three subjects for ten consecutive periods.

The public goods game consists of a contribution stage in which each subject  $i \in \{1, 2, 3\}$  receives an endowment of 20 tokens. Subjects simultaneously decide how many tokens,  $c_i \in [0, 20]$ , they contribute for the provision of the public good. The marginal monetary benefit for subject  $i$  for every token contributed by any group member is given by  $\alpha_i$  whereas the marginal monetary benefit from not contributing a token equals one. Hence, from a purely monetary perspective, if  $\alpha_i < 1$ , then subject  $i$ 's dominant strategy is to contribute nothing because any contribution reduces own earnings. If, however,  $\alpha_i > 1$ , then subject  $i$ 's dominant strategy is to contribute the whole endowment independent of what the other group members contribute. Additionally, if  $\sum_i \alpha_i > 1$ , then each token contributed increases the sum of earnings in the group and, hence, monetary surplus maximization is attained if everyone contributes all 20 tokens.<sup>6</sup>

---

<sup>6</sup>In the experiment all subjects are informed that they will interact for exactly ten rounds with the same individuals. Furthermore, they know the value of  $\alpha_i$  of all group members but neither during nor after the experiment do they get to know their identities.



In the punishment condition the contribution stage of the public goods game is followed by a punishment stage. In the punishment stage subjects are first informed of the contributions of all other group members. Thereafter, each subject  $i$  simultaneously decides how many punishment points,  $p_{ij} \in [0, 10]$ , to assign to each other subject  $j$ . Each punishment point costs the punisher one token and reduces the earnings of the punished subject by three tokens.<sup>7</sup> After the punishment decision subjects' are informed of the total number of punishment points assigned to them by other subjects in their group. As in Fehr and Gächter (2000) subjects receive no information about individual punishment behavior of other group members. At the end of a period, earnings of a subject  $i$  are given by:<sup>8</sup>

$$\pi_i = 20 - c_i + \alpha_i \sum_j c_j - 3 \sum_{j \neq i} p_{ji} - \sum_{j \neq i} p_{ij}.$$

Importantly, in some treatments groups differ in regard to the values of marginal benefits,  $\alpha_i$ , assigned to subjects in a group. In *normal* groups all subjects have the same  $\alpha_i = 0.5$ . In *privileged* groups, at the beginning of the experiment a subject is randomly chosen to receive an  $\alpha_j = 1.5$ . The two other subjects in a privileged group have the same  $\alpha_i = 0.5$ . Each subject keeps his or her assigned marginal benefit for all ten periods. For convenience, we refer to subjects with an  $\alpha_i = 0.5$  as *low-benefit* subjects and to subjects with an  $\alpha_j = 1.5$  as *high-benefit* subjects. Clearly, from a monetary perspective high-benefit subjects have a dominant strategy to contribute all their endowment. Note, that the presence of a high-benefit subject precisely mirrors Olson's definition of privileged groups as being groups where at least one member is willing to bear the cost of providing some level of the public good on his own. Note that, due to the presence of low-benefit subjects, privileged groups are also prone to under-provision of the public good. The four different treatments are summarized in Table 1.

Normal and privileged groups differ only in the absence or presence of a high-benefit subject instead of a low-benefit subject. Thus, by comparing these group types we can investigate the effect of the high-benefit subject on overall contribution levels when punishment is or is not available. Furthermore, since low-benefit subjects face the same monetary incentives in both types of groups, any differences in behavior of these subjects across group types must be due to the presence of the high-benefit subject.

<sup>7</sup>As it is standard in experiments of public goods games with punishment, subjects cannot be punished below zero tokens but can incur losses through the costs of punishing others.

<sup>8</sup>In case a subject makes losses these are calculated as follows:  $\pi_i = \max[0, 20 - c_i + \alpha_i \sum_j c_j - 3 \sum_{j \neq i} p_{ji}] - \sum_{j \neq i} p_{ij}$ .

Table 1: Experimental treatments

	Baseline	Punishment
Normal Groups	Only contribution stage $\alpha_i = 0.5$ for all subjects	Contribution and punishment stages $\alpha_i = 0.5$ for all subjects
Privileged Groups	Only contribution stage $\alpha_j = 1.5$ for one subject, $\alpha_i = 0.5$ for others	Contribution and punishment stages $\alpha_j = 1.5$ for one subject, $\alpha_i = 0.5$ for others

Under the assumption that individuals are selfish money maximizers, nobody is predicted to contribute a positive amount to the public good in normal groups. However, the evidence from previous public goods experiments suggests some voluntary contribution (see e.g., Ledyard, 1995; Zelmer, 2003). The typical contribution pattern found in experimental studies is that subjects contribute substantial positive amounts in early periods and that contribution levels decline to low levels with repetition. Such a pattern is also to be expected in our treatment with normal groups without punishment. In privileged groups in the baseline treatment, low-benefit individuals face the same monetary incentives as subjects in the normal groups. In contrast, high-benefit individuals earn more money the more they contribute to the public good. Therefore, when taking only material incentives into account, it is predicted that public good provision is higher in privileged groups than in normal groups and that high-benefit subjects contribute all their endowment whereas low-benefit subjects contribute nothing. Note, however, that the simultaneous presence of high- and low-benefit subjects may interact with non-monetary motives of voluntary contribution in a non-trivial way. This is shortly discussed below.

Since punishment is costly and hence not credible for rational own-payoff maximizing individuals, predictions based on this assumption are the same as without a punishment option. However, there is considerable evidence that punishment is used to enforce high and stable contribution levels (Fehr and Gächter, 2000). Various models of generalized preferences have been proposed to explain contribution and punishment behavior in a range of experiments.<sup>9</sup> In the following we briefly discuss some general predictions regarding punishment and voluntary contributions of such models and relate it to our experimental set-up with privileged groups.

Outcome based models of inequity aversion predict that in normal groups contribution levels will be relatively low in the baseline treatment and can be high in the punishment treatment (see

<sup>9</sup>For example, Levin (1998), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002), Falk and Fischbacher (2006).

e.g. Fehr and Schmidt, 1999). For privileged groups the predictions are different. Consider the situation without punishment first. Any contribution of the high-benefit individual increases his income and does not affect income differences, because the marginal net gain of each individual is 0.5. Hence, high-benefit individuals contribute all their endowment irrespective of their other-regarding preferences. In contrast, any contribution by low-benefit individuals decreases their income and increases income differences. Consequently, low-benefit individuals have neither a material nor an inequity based incentive to contribute. Introducing the punishment opportunity does not change this prediction. To see this, consider that the high-benefit individual contributes the whole endowment and the low-benefit individuals contribute nothing. In this case everyone in the group earns 30 tokens. Since the motivation for punishment is to decrease income differences and since any contribution by a low-benefit individual increases inequality, there are no incentives under inequity aversion for a high-benefit individual to force low-benefit group members to contribute.<sup>10</sup> Hence, in contrast to normal groups, the opportunity to punish is not predicted to increase contributions in privileged groups.

Predictions of intention-based models of social preferences are more ambiguous. In normal groups, high and low contributions have a straightforward interpretation as being a kind or an unkind action. This provides an incentive to punish low contributors and to not punish high contributors. In privileged groups, however, it is difficult to judge the kindness of actions of high-benefit individuals. In particular, high contributions directly benefit such individuals and hence may be interpreted as a selfish act. Consequently, in intention based models the prediction for privileged groups depends critically on how intentions are specifically modeled. For example, in the model of Falk and Fischbacher (2006), where kindness depends on relative income positions, low-benefit individuals will not regard as kind the high contributions by high-benefit individuals, and high-benefit individuals will not consider low contributions by low-benefit individuals as unkind and thus deserving of punishment.

In summary, purely outcome-based as well as intention-based models of social preferences suggest that introducing punishment opportunities has a weaker effect on contributions in privileged groups than in normal groups.

---

<sup>10</sup>Certainly, as long as earnings end up equal, punishment can be used to enforce other cooperation levels. It would imply that, in equilibrium, low-benefit individuals punish high-benefit individuals simply to reduce income differences. However, all such equilibria are Pareto-dominated by the equilibrium where low-benefit individuals do not contribute.

### 3 Results

The computerized experiment was conducted at the CREED laboratory of the University of Amsterdam. The experiment was programmed with z-Tree Fischbacher (2007). In total 81 subjects participated in the experiment. 21 (18) participated in the baseline treatment without (with) punishment and 24 (18) in the privileged group treatment without (with) punishment. This generated between 6 and 8 independent observations per treatment. Each treatment was run in a separate session. About half of the subjects were female and also around half were students of economics (the other half came from various fields such as biology, engineering, political science, and law). A typical experimental session took one hour. On average, subjects were paid out €13.76 (approx. US\$17.50). More detailed experimental procedures and the instructions can be found in Appendix A. Detailed descriptive statistics of the data are provided in Appendix B.

#### 3.1 Voluntary Contributions

In this section we first discuss contribution behavior at the group level. Thereafter, we zoom into the behavior of low- and high-benefit subjects.

Figure 1 depicts the average contribution levels over periods in (a) the baseline treatments and (b) the punishment treatments for both group types. In keeping with evidence from comparable public goods experiments, average contributions to the public good are relatively low and decrease over time in the baseline treatment with normal groups. At a somewhat higher level a similar dynamics of contributions is observed in privileged groups. On average normal groups contribute 4.21 tokens whereas privileged groups contribute 8.68 tokens. A non-parametric Mann-Whitney test clearly rejects equality of distributions ( $p = 0.001$ ).<sup>11</sup>

In the baseline treatment, rank order correlations (Spearman's  $\rho$ ) of contributions on periods indicate that voluntary contributions significantly decline over time in normal as well as privileged groups ( $\rho < -0.462$ ,  $p < 0.001$ ). However, compared to privileged groups, contributions decline to a lower level in normal groups. This difference may be attributed to convergence towards different equilibrium contribution levels. Indeed, in normal groups, we observe a decline

---

<sup>11</sup>Unless otherwise noted, for between-treatment comparisons we use a Mann-Whitney test, which tests the hypothesis that two independent samples are drawn from the same population. For within-treatment comparisons we use the Wilcoxon signed-ranks test for matched pairs. For a description of all tests used in this paper see Siegel and Castellan (1988). If it is not otherwise indicated we apply one-sided test statistics and use group averages across all periods as independent observations. All reported  $p$ -values are exact values.

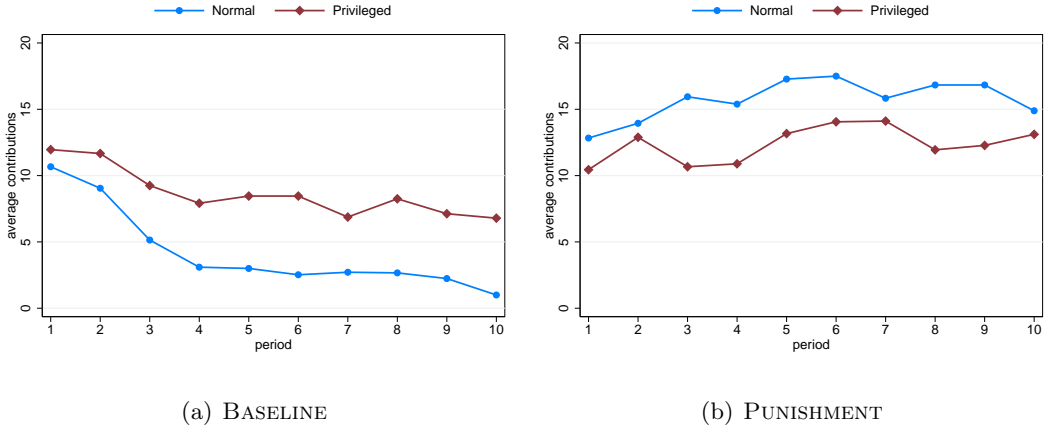


Figure 1: Contributions in the different groups and treatments

to (almost) zero contributions. In privileged groups last period contributions are on average 6.79 tokens. This is only slightly above the 6.67 tokens, obtained when the high-benefit subject contributes everything and the two low-benefit subjects do not contribute at all.

In the punishment treatment contribution levels are higher than in the baseline treatment and they do not decline over time for both group types. Note that, in contrast to the baseline without punishment, it is no longer the case that contributions are highest in privileged groups. In fact, the average contribution in normal groups is 15.73 whereas it is only 12.36 in privileged groups. Although, a Mann-Whitney test does not reject the hypothesis that average contributions come from the same distribution ( $p = 0.120$ ) it is intriguing that privileged groups lose their privileged status once punishment is possible. Additionally, according to Spearman's rank order correlations, contributions stay constant over time in privileged groups ( $\rho = 0.149$ ,  $p = 0.129$ ) but exhibit a significant upward trend in normal groups ( $\rho = 0.248$ ,  $p = 0.029$ ).

The quantitative effect of punishment on contributions also strongly differs between group types. In normal groups the increase in contributions due to punishment is large (11.52 tokens) and statistically significant ( $p < 0.001$ ). For privileged groups, the increase is also statistically significant ( $p = 0.041$ ) but quantitatively rather small (3.68 tokens). The difference in the effectiveness of punishment is also illustrated by the fact that in normal groups full contribution by all group members is achieved in 35.5 percent of all periods whereas in privileged groups it is observed only in 6.7 percent. Hence, although the opportunity to punish increases contributions in both types of groups the magnitude of this effect is strikingly different between normal and privileged groups. Particularly, it is much stronger in groups where all subjects benefit equally from the public good. We summarize the findings discussed so far in our first result.

**Result 1** VOLUNTARY CONTRIBUTIONS IN NORMAL AND PRIVILEGED GROUPS.

*In the absence of punishment, there is under-provision of the public good in normal as well as privileged groups, although voluntary contributions are higher in the latter. If punishment is possible, there is still under-provision but at a lower level and the difference between group types disappears. Punishment possibilities increase contributions in both group types but its effect is much smaller in privileged groups.*

Given the diametrically opposed material incentives for high- and low-benefit subjects it is to be expected that their contribution behavior differs. Figure 2 shows that this is indeed the case. In privileged groups, high-benefit subjects contribute significantly more than low-benefit subjects in treatments with and without punishment. For high- and low-benefit subjects the average contribution levels are 17.03 and 4.50, respectively, in the baseline treatment, and 18.43 and 9.32, respectively, in the punishment treatment. In both cases equality of distribution is rejected at any conventional level of statistical significance ( $p < 0.005$ ).

Remarkably, high-benefit subjects do not follow their dominant strategy to contribute the full endowment of 20 tokens. This behavior is particularly pronounced and statistically significant in the baseline treatment without punishment ( $p = 0.010$ ). Recall that for high-benefit subjects contributing the whole endowment is a dominant strategy in absolute as well as relative material terms, and it is also the best for their group's total payoff. Therefore, it is impossible to explain the observed deviations from full contribution on the basis of material payoffs only. High-benefit subjects act (at least to some extent) conditionally cooperative in terms of contributions. It appears that they “punish” low-benefit subjects for low contributions by not contributing all their endowment. Three observations provide further supporting evidence for this interpretation. First, albeit contributing everything is a dominant strategy Figure 2(a) shows that contributions of high-benefit subjects decrease over time, until in period 8 an end-game effect sets in. For periods 1-7, a Spearman rank order correlation weakly corroborates the visual impression ( $\rho = -0.175$ ,  $p = 0.098$ ). Second, periods where the high-benefit subject does not fully contribute are preceded by periods where low-benefit subjects contribute significantly less than in periods preceding full contribution by the high-benefit subject ( $p = 0.034$ ). This behavior is clearly consistent with a tit-for-tat-like punishment strategy. Third, in the treatment with the explicit punishment possibility under-provision of high-benefit subjects is less pronounced than in the baseline treatment. Indeed, when testing if actual contributions of high-benefit subjects are equal to full contribution, a Wilcoxon signed-rank test clearly rejects equality for the baseline treatment ( $p = 0.010$ ) but only marginally for the punishment treatment ( $p = 0.080$ ). This difference in contribution behavior is consistent with the interpretation

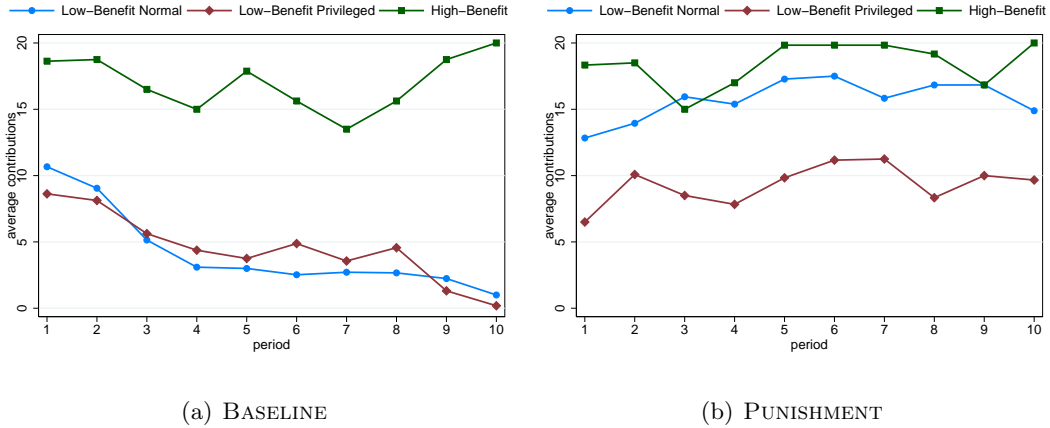


Figure 2: Contributions by low- and high-benefit subjects in each treatment

that in the presence of the more effective explicit punishment possibility there is no need for high-benefit subjects to resort to lowering contributions as a means of disciplining low-benefit subjects.<sup>12</sup> In the next section we provide evidence showing that high-benefit subjects indeed do punish low-benefit subjects.

In the no-punishment treatment, low-benefit subjects behave very similar in normal and in privileged groups (see Figure 2(a)). Average contributions in both types of groups are almost equal (4.21 and 4.50 in normal and privileged groups, respectively) and statistically indistinguishable ( $p = 0.347$ ). Contributions of low-benefit subjects also decline significantly over time in both types of groups (Spearman's  $\rho = -0.584$  and  $\rho = -0.590$ ,  $p < 0.001$ ). In contrast, in the punishment treatment, contributions of low-benefit subjects differ starkly in normal and in privileged groups (see Figure 2(b)). Average contributions amount to 15.73 tokens in normal groups but are only 9.32 tokens in privileged groups. A Mann-Whitney test clearly rejects equality of distributions ( $p = 0.033$ ). Interestingly, this difference in contributions is present already in the first period. In period 1, the average contribution in normal groups is 12.83, which is almost twice as high as in privileged groups where it is only 6.50 tokens. Statistically, equality of contributions is clearly rejected ( $p = 0.004$ ). Moreover, this difference does not vanish but rather increases over time. Spearman rank order correlations show that in normal groups contributions exhibit a statistically significant upward trend ( $\rho = 0.247$ ,  $p = 0.028$ ) whereas no such trend can be detected in privileged groups ( $\rho = 0.104$ ,  $p = 0.215$ ).

<sup>12</sup>Another possibility for under-provision is spiteful behavior as reported by Saijo and Nakamura (1995) in groups where  $\alpha_i > 1$  for all group members. However, this fails to explain the end-game effect in the baseline treatment and the observed relation with the contribution behavior of low-benefit subjects. Brandts and Schram (2001) also find a small degree of under-provision, which they attribute to errors. However, they do not explore the conditional cooperation possibility.

Clearly, in privileged groups contributions of low-benefit subjects are significantly higher in the punishment treatment than in the baseline treatment ( $p = 0.041$ ). However, in comparison to the treatment effect for normal groups the observed increase in contributions is rather meager. For normal groups, the average increase in contributions from the baseline treatment to the punishment treatment is 11.52 tokens, whereas it is only 4.82 tokens for low-benefit subjects in privileged groups ( $p = 0.021$ ). Together these observations indicate that the effectiveness of punishment concerning low-benefit subjects strongly differs between group types to the disadvantage of privileged groups. We summarize our findings in the following result.

**Result 2** CONTRIBUTIONS OF HIGH- AND LOW-BENEFIT SUBJECTS.

*Without punishment opportunities, although high-benefit subjects contribute significantly more than low-benefit subjects there is significant under-provision of the public good by high-benefit subjects. With punishment opportunities, contributions of both types of subjects increase. However, contributions of low-benefit subjects in privileged groups reach only 60 percent of the contribution level in normal groups.*

### 3.2 Punishment

The amount of tokens lost, per subject and period, due to punishment is similar in privileged and normal groups (3.72 and 3.23,  $p = 0.409$ ). In privileged groups low-benefit subjects are punished slightly more than high-benefit subjects (4.03 vs. 3.10). However, this difference is not statistically significant ( $p = 0.310$ ). Over time punishment decreases significantly in both types of groups (Spearman's  $\rho < -0.365$ ,  $p < 0.005$ ). Allocated punishment points are also similar across types. On average, low-benefit subjects in normal and in privileged groups deal out 0.54 and 0.51 punishment points, respectively. High-benefit subjects in privileged groups assign, with 0.84, slightly more punishment points. The differences are statistically not significant ( $p > 0.155$ ).

In the remainder of this subsection we first discuss possible determinants of punishment and how they differ, first, between group types and, second, between high- and low-benefit subjects in privileged groups. Thereafter, we investigate if and how high- and low-benefit subjects react with their public good contributions to received punishment. This will shed light on the observed differences in contributions of low-benefit subjects in normal and privileged groups.

For the analysis of possible determinants of punishment behavior we apply Tobit regressions. The independent variable is the amount of punishment points,  $p_{ij}$ , subject  $i$  allocates to subject



$j$ .<sup>13</sup> As explanatory variables we use the deviation of  $j$ 's contribution from  $i$ 's contribution, and the deviation of  $j$ 's contribution from the contribution of the third subject in the group  $k$ . This provides a more complete picture of the dependence of punishment on relative contributions than the deviation from average group contribution used in some other studies. Another reason is that in privileged groups the deviation from the average may not form a meaningful focal point from which to judge voluntary contributions for both high- and low-benefit subjects. Results from other studies with only normal groups and intuition suggests that negative deviations are more heavily punished than positive differences (Fehr and Gächter, 2000). We, therefore, allow for different coefficients for positive and negative deviations. In addition, we control for  $j$ 's total contribution, for cases in which all subjects contributed their whole endowment (this accounts for 21% of all periods but only 9% of all punishment points), and for a time trend. For privileged groups, we run separate regressions for high- and low-benefit subjects. In the regressions for low-benefit subjects we included a dummy variable if the target person  $j$  is a high-benefit subject and also interact this variable with the period of play.<sup>14</sup> The results are reported in Table 2.

We first discuss behavior of low-benefit subjects in normal and privileged groups (columns two and three in the table). In line with other studies, in both groups negative deviations from own contributions are strongly and statistically significantly punished. In this respect, there is no statistically significant difference between normal and privileged groups (Wald test,  $p = 0.982$ ). In normal groups, positive deviations from own contributions are also punished, albeit at a much lower level than negative deviations. Such 'perverse' punishment of high contributors is also found by other authors (see e.g., Egas and Riedl, 2005; Cinyabuguma et al., 2006; Gächter and Herrmann, 2006). In privileged groups such punishment is statistically not significant. Probably the most interesting difference between groups is the punishment response towards differences in contributions between the target person  $j$  and the third person  $k$ . In normal groups, a positive deviation of  $j$  compared to  $k$  leads to a strong and significant decrease in punishment whereas in privileged groups no such effect is detected (Wald test for the difference between treatments,  $p = 0.004$ ). In privileged groups the highest contributor is almost always the high-benefit subject. Therefore, the above result indicates that relatively

---

<sup>13</sup>The independent variable,  $p_{ij}$ , is censored below as well as above. Tobit estimates take this into account. All reported standard errors are based on robust estimates and dependence of observations within groups is accounted for by clustering on each period and group.

<sup>14</sup>If we drop these two explanatory variables concerning the high-benefit subject all reported results stay quantitatively similar and qualitatively the same. The regression coefficients of the other explanatory variables change only marginally and all reported (non-)significances remain.

Table 2: Punishment assigned to  $j$  by  $i$  depending on differences in contributions

Dependent variable: $p_{ij}$	<i>Normal Groups</i>		<i>Privileged Groups</i>	
	Punishment given by $i$ to $j$		All	Low-benefit
Contribution of $j$		0.121	0.120	-0.086
$c_j$		(0.077)	(0.095)	(0.117)
Positive deviation from $c_i$		0.148***	-0.105	-0.311
$\max(c_j - c_i, 0)$		(0.057)	(0.113)	(0.333)
Negative deviation from $c_i$		0.312***	0.420***	-0.074
$\max(c_i - c_j, 0)$		(0.071)	(0.135)	(0.102)
Positive deviation from $c_k$		-0.537***	-0.137	-0.166
$\max(c_j - c_k, 0)$		(0.143)	(0.113)	(0.109)
Negative deviation from $c_k$		0.118	0.058	0.132
$\max(c_k - c_j, 0)$		(0.091)	(0.106)	(0.136)
Fully contributing group		-4.360***	2.977	-14.606***
1 if $c_i + c_j + c_k = 60$		(0.921)	(2.534)	(2.055)
Period		-0.081	-0.019	-0.457**
$t$		(0.071)	(0.180)	(0.188)
$j$ is a high-benefit subject			5.000***	
1 if $\alpha_j = 1.5$			(1.474)	
$j$ is a high-benefit subject $\times$ Period			-0.696***	
$t$ if $\alpha_j = 1.5$			(0.189)	
Constant		-3.253**	-6.292***	2.721
		(1.496)	(2.419)	(2.237)
# Observations		360	240	120
Log-likelihood		-271.964	-191.053	-140.377
$\chi^2$		124.88	38.65	56.71

Note: Numbers between brackets are standard errors \*\*\* Significant at the 1% level.

\*\* Significant at the 5% level. \* Significant at the 10% level.

high contributions of such a subject are not rewarded by less punishment. Additionally, low-benefit subjects appear to punish high-benefit subjects simply for being a high-benefit subject, especially at the beginning of the game (see Table 2, column three).

In contrast to low-benefit subjects, punishment behavior of high-benefit subjects is largely independent of relative contributions except that they reward full contributions of low-benefit subjects with a strong and significant decrease in punishment. High-benefit subjects also punish less in later periods as indicated by the statistically significantly negative linear time trend. Recall, that on average high-benefit subjects do not punish less than low-benefit subjects.

Therefore, it is the lack of sensitivity of high-benefit subjects regarding relative contributions of low-benefit subjects that might account for the comparatively low contributions of low-benefit subjects in privileged groups.

**Result 3** PUNISHMENT BEHAVIOR IN NORMAL AND PRIVILEGED GROUPS.

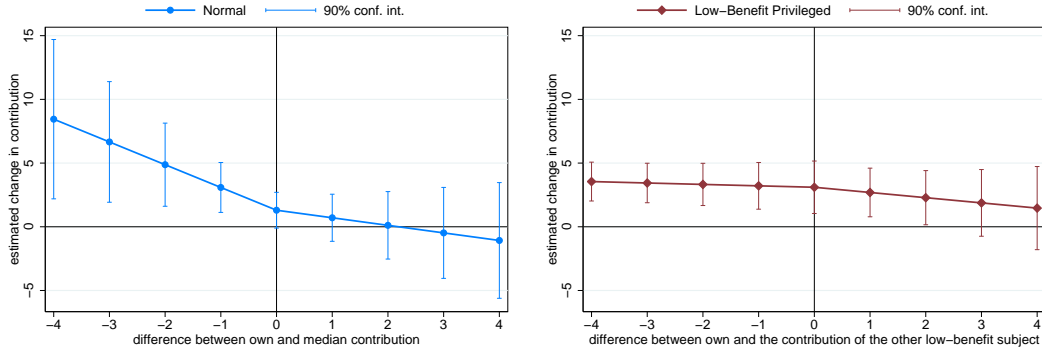
*In both types of groups low-benefit subjects' punishment behavior is sensitive to relative contributions. They severely punish deviations from their own contribution. Additionally, in normal groups low-benefit subjects punish less if the target person contributes more than the third group member. In contrast, high-benefit subjects' punishment behavior is insensitive to the relative contributions of low-benefit subjects.*

The effectiveness of punishment in increasing contributions to the public good crucially depends on the response of punished players. To investigate this response and its potentially different characteristics across group types we have calculated the effect of being punished (versus not being punished) in a period  $t - 1$  on the change of contributions from period  $t - 1$  to period  $t$ . We denote this change of contributions by  $\Delta c_i$ . The reported results are based on OLS regressions which can be found in Appendix C. Figure 3 shows this estimated response for low-benefit subjects in both types of groups for selected deviations in contributions. For normal groups (panel (a)) it shows the effect of being punished on  $\Delta c_i$  for deviations between +4 and -4 points in comparison to the median contributor in the group. For low-benefit subjects in privileged groups (panel (b)) it shows the effect for deviations between +4 and -4 points in comparison to the other low-benefit subject.<sup>15</sup> In privileged groups in 90.0% of all cases a low-benefit subject is indeed the median contributor. Therefore, a comparison between normal and privileged groups regarding the effect of punishment on  $\Delta c_i$  for low-benefit subjects is informative, at least for negative deviations.

In normal groups, punishment of subjects who contributed the same or more than the median contributor has no statistically significant effect on contributions (Figure 3(a)). Interestingly, in privileged groups, this only holds for punished subjects who contributed at least 3 points more than the other low-benefit subject, whereas those who contribute the same, 1, or 2 points more increase their contributions significantly (Figure 3(b)). This difference in response of low-benefit subjects in normal and privileged groups is consistent with their different relative positions in the two group types. In normal groups the subjects in question are the highest

---

<sup>15</sup>In the estimation for privileged groups we assume a negative deviation of 10 points compared to the high-benefit subject. Since this is the actual median deviation of low-benefit subjects from high-benefit subjects it is a reasonable assumption.



(a) NORMAL GROUPS

(b) PRIVILEGED GROUPS

Figure 3: Effect of being punished on change in contribution for low-benefit subjects

contributors whereas they are only median contributors, usually contributing much less than the high-benefit subject, in privileged groups.

Low-benefit subjects who are the lowest contributors in normal as well as in privileged groups respond with statistically significant increases in contributions. If contributions are close to the median contribution (around  $-1$  point) the effect is quantitatively similar in both group types. However, for larger negative deviations there are interesting and pronounced differences. For instance, on average, for negative deviations of 4 points, punishment increases contributions about twice as much in normal groups as in privileged groups. This is because more free-riding by low-benefit subjects (i.e. a larger negative difference with the median contributor) produces a much weaker positive response in contributions in privileged groups than in normal groups. This can be seen in Figure 3 as a steeper slope for the estimated  $\Delta c_i$  for negative deviations from the median contribution. A Wald test confirms that the slope of normal groups is significantly steeper than that of privileged groups ( $p = 0.044$ ). We summarize our findings in the following result.

**Result 4** RESPONSE TO PUNISHMENT IN NORMAL AND PRIVILEGED GROUPS.

*In both types of groups, low-benefit subjects who are the lowest contributors respond with a significant increase in contributions if punished. For strong free-riders this effect is stronger in normal groups. Low-benefit subjects who deviate positively from the other low-benefit subjects do not respond positively to punishment in normal groups and only slightly so in privileged groups.*

Recall that high-benefit subjects in privileged groups exhibit a relatively unstructured punishment behavior (Result 3). This fact together with the above finding, that strongly free-riding

low-benefit subjects exhibit a relatively weak response to punishment in privileged groups, explains why the opportunity to punish has a much smaller effect on the contributions of low-benefit subjects in privileged groups compared to normal groups (Result 2).

### 3.3 Efficiency and Inequality

The socially beneficial effect of the opportunity to punish free-riders is that it increases contributions to the public good and thereby also total monetary earnings. The downside is that if costly punishment is actually administered it destroys resources and hence decreases total monetary surplus. In this section we examine how punishment affects earnings of low- and high-benefit subjects in normal and privileged groups.

Taken across all periods, in the baseline treatment with normal groups average earnings per subject and per period are equal to 22.10 tokens which is statistically not significantly different from the 23.67 average earnings when punishment is possible ( $p = 0.366$ ). Inspecting the dynamics of earnings, however, reveals a clear divergence between the baseline and the punishment treatment. In the baseline treatment earnings significantly decrease over time whereas they show a clear upwards trend in the punishment treatment (Spearman's  $\rho = -0.584$ ,  $p < 0.001$  and  $\rho = 0.322$ ,  $p = 0.006$ ). Furthermore, earnings in the second half (periods 6-10) of the public goods game are significantly larger in the punishment treatment compared to the baseline treatment (21.11 tokens and 24.86 tokens,  $p = 0.017$ ).

In privileged groups the effect of punishment on earnings is more ambiguous. As in normal groups, when taken across all periods, average earnings per subject and period are similar in treatments without and with punishment (33.01 and 33.58 tokens,  $p = 0.426$ ). Earnings also show a decreasing trend in the baseline treatment and an increasing trend in the punishment treatment (Spearman's  $\rho = -0.463$ ,  $p < 0.001$  and  $\rho = 0.254$ ,  $p = 0.025$ ). However, unlike in normal groups, in later periods (6-10) the earnings in the punishment treatment are only marginally significantly larger than in the baseline treatment (31.25 tokens without punishment and 36.23 tokens with punishment,  $p = 0.091$ ). The reason for this difference between normal and privileged groups is that in privileged groups, low-benefit subjects increase their contributions to a lower extent in the punishment treatment.

Punishment not only affects total earnings but it may also have an effect on how these earnings are distributed. A natural and simple measure of inequality is the standard deviation of earnings per period within each group. In normal groups the average standard deviation across periods is 4.76 in the baseline treatment and 3.43 when punishment is possible. Since a

Mann-Whitney test does not reject equality of distributions it appears that earnings are similarly distributed in both treatments ( $p = 0.183$ ). In privileged groups, however, punishment strongly increases inequality. Whereas in the baseline treatment the average standard deviation is 11.56 it increases to 23.69 in the punishment treatment. The difference in distributions is statistically significant ( $p = 0.041$ ).

The underlying reason for increased inequality in privileged groups is that punishment has a stark differential effect on the earnings of low- and high-benefit subjects. High-benefit subjects gain from the punishment institution: their earnings increase from 42.01 tokens in the baseline treatment to 52.38 tokens in the punishment treatment ( $p = 0.071$ ). In contrast, low-benefit subjects in privileged groups do not gain at all from the punishment institution. In effect, their losses due to punishment result in significantly lower earnings in the punishment treatment compared to the baseline treatment (28.51 tokens without punishment and 24.18 tokens with punishment,  $p = 0.004$ ). If we compare earnings of low-benefit subjects in normal and privileged groups we find that in the treatment without punishment low-benefit subjects earn significantly more in privileged groups ( $p < 0.001$ ), but this is no longer the case in the treatment with punishment ( $p = 0.409$ ). In other words, low-benefit subjects ‘exploit’ the high-benefit subject when punishment is not possible, but are no longer able to do so when punishment is introduced. This is important as it illustrates that privileged groups are not necessarily characterized by the exploitation of the ‘rich by the poor’ as is usually claimed (Olson, 1965; Sandler, 1992). We summarize the findings of this section in our final result.

**Result 5** EFFICIENCY AND INEQUALITY.

*In normal groups, the opportunity to punish leads to higher overall earnings with no effect on their distribution. In privileged groups, punishment does not significantly increase total earnings but strongly increases inequality at the cost of low-benefit subjects. Consequently, with punishment, low-benefit subjects do not benefit from being part of a privileged group.*

## 4 Discussion and Conclusion

In this paper we provide evidence that Olson’s (Olson, 1965) privileged groups are not as privileged as conjectured, especially if punishment is possible. In line with standard theoretical predictions contributions are higher in privileged groups than in normal, non-privileged, groups and under-provision of the public good is observed in both group types. Surprisingly, in privileged groups this is not only due to free-riding by low-benefit subjects but also due

to non-optimal contributions of subjects whose dominant strategy is to contribute fully. We observe that high-benefit subjects reciprocate to low contributions of low-benefit subjects by lowering their contributions, despite the fact that contributing fully maximizes absolute and relative individual earnings as well as group earnings. When a punishment option is introduced privileged groups lose their privileged status completely. Total contributions in normal and privileged groups are virtually identical. In addition, the introduction of punishment significantly increases earnings inequality in privileged groups as high-benefit subjects avoid being exploited by low-benefit subjects. We identify the following behavioral features that account for this relative disadvantage of privileged groups. With a punishment option, low-benefit subjects in privileged groups increase their contributions to a much smaller extent than their counterparts in non-privileged groups where all have the same incentive to free-ride. The reason for this is twofold. First, high-benefit subjects exhibit an unstructured punishment pattern and, second, low-benefit subjects increase their contributions in response to punishment to a much smaller extent in privileged groups than in normal groups.

Existing models of other-regarding preferences capture to some extent the possibility that punishment is less effective in privileged groups than in normal groups. The theoretical argument supporting this result is not in agreement with our empirical observations, however. In these models low-benefit subjects do not contribute because high-benefit subjects are unwilling to punish them since doing so would increase earnings disparities. What we find is that high-benefit subjects do punish low-benefit subjects who free-ride (albeit in a relatively unstructured way), but that low-benefit subjects are reluctant to increase contributions despite the fact that they are punished. This is an interesting finding because it demonstrates that subjects in one role (high-benefit) may be willing to undertake a costly action (punishing) although it increases earnings inequalities but that subjects in another role (low-benefit) are unwilling to take a costly action (increasing contributions) because it increases inequality.

Our empirical results highlight the importance of a common notion of ‘fair’ or ‘kind’ behavior for mechanisms like decentralized punishment to be effective. In homogeneous groups contributions to a public good are unambiguously kind because it increases everyone’s payoff at a cost for the contributing individual. In privileged groups, however, high contributions by high-benefit subjects may be perceived as selfish (because they increase the contributor’s earnings), unfair (because they increase earning inequalities), or kind (because they increase the earnings of others). Low contributions of low-benefit subjects can similarly be interpreted in different ways. The fact that high-benefit subjects punish low-benefit subjects but low-benefit subjects only reluctantly increase contributions is consistent with self-serving interpretations

of the ‘correct’ fairness norm (see e.g. Babcock et al., 1995; Gächter and Riedl, 2005). That diverging perceptions of what constitutes fair behavior may have detrimental effects on cooperation has also been suggested in recent research on emotions and social behavior. For instance, Hopfensitz and Reuben (2005) and Reuben and van Winden (2005) find that sanctions enforce cooperative behavior only if people feel they have transgressed a social norm and hence deserve punishment (see also Bowles and Gintis, 2005). If low-benefit subjects in privileged groups conceive that they do not deserve sanctioning, a weak response as observed in our experiment may be the consequence. In summary, in privileged groups different perceptions of fair behavior are the likely reason for costly punishment being relatively ineffective and hence wasteful.

In an early study, Johnson and Libecap (1982) argue theoretically that group heterogeneity considerably complicates the private enforcement of cooperative behavior. The results of our paper add to this finding the experimental evidence that the behavioral effectiveness of an institution in overcoming social dilemmas can crucially depend on group characteristics. An observation also in line with results showing that sanctioning institutions can have very different effects on participants behavior with diverging cultural backgrounds (Gächter and Herrmann, 2006). This implies that the design of effective enforcement institutions, formal as well as informal ones, asks for a proper understanding of the interplay of heterogeneous group characteristics, as in privileged groups, behavioral inclinations and biases, and incentives.



# A Experimental Procedures and Instructions

## Experimental procedures

After arrival in the lab's reception room, each subject drew a card to be randomly assigned to a seat in the laboratory. Once everyone was seated, the instructions for the experiment were read aloud (a translation of the instructions is provided below). In the privileged treatments (with and without punishment), subjects were informed in the instructions whether they would be a high-benefit subject or a low-benefit subject. Thereafter, subjects had to answer a few exercises in order to check their understanding of the game. Next, the subjects played the repeated public goods game with or without punishment via the computer. Once the game ended subjects answered a debriefing questionnaire after which they were paid in private and dismissed.

## Instructions

These are the instructions given to high-benefit subjects in the privileged treatment. The instructions given to low-benefit subjects and to subjects in the baseline treatment are available upon request.

## Introduction

This experiment is divided into different periods. There will be 10 periods in total. During all 10 periods, the participants are divided into groups of three. Therefore, you will be in a group with 2 other participants. The composition of the groups will remain the same during all of the experiment.

Each period consists of two stages. In the first stage, you have to decide how many tokens you contribute to a group project. In the second stage, you will learn how much the other members of your group contributed to the project.

## The first stage

At the beginning of each period each participant in your group receives 20 tokens. We will refer to these tokens as the initial endowment.

In the first stage you decide how to use your initial endowment. You have to choose how many tokens you want to contribute to a group project and how many of them to keep for yourself. You can contribute any amount of your initial endowment to the group project. How many tokens you contribute is up to you. Each other group member will also make such a decision. All decisions are made simultaneously. That is, nobody will be informed about the decision of the other group members before everyone made his or her decision.

## Earnings in the first stage

Your earnings in tokens, in each period, are the sum of two parts:

- The number of tokens that you kept for yourself.
- Your income from the group project. This income equals:

$$[\text{multiplication factor}] \times \text{sum of contributions of all group members to the project}$$

The projects' multiplication factor is determined as follows: in each group, *one of the group members will have a multiplication factor of 1.5* and the other *two group members will have a multiplication factor of 0.5*. Before the experiment started each desk was assigned a multiplication factor equal to either 0.5 or 1.5. Therefore, by randomly assigning the yellow cards, each participant was randomly assigned to one of these values. The multiplication factor will be the same for all the 10 periods. *You will be the group member who has a multiplication factor of 1.5.*

Notice that, for each token which you keep for yourself you earn 1 token. If instead you contribute this token to the group project, then the total contribution to the project will rise by one token. Your income from the group project will rise by 1.5 tokens. Moreover, the other group members' income from the project will rise by 0.5 tokens. Your contribution to the group project therefore also raises the income of the other group members. For each token contributed to the project the total earnings of the group will rise by 2.5 tokens. Note that, you also earn tokens for each token contributed to the group project by the other group members. For each token contributed by any member you earn 1.5 tokens.

In summary, your earnings in tokens at the first stage of a period are equal to:

$$20 - \text{your contribution} + \text{your multiplication factor} \times (\text{sum of contributions})$$

After everyone has made his or her decision the first stage ends.

### **Example for the first stage**

Here is an example that illustrates how the earnings in tokens are calculated in the first stage of each period. The numbers used in the example are arbitrarily chosen.

You are in a group with two other participants (group member 1 and group member 2). Each participant's multiplication factor equals: you = 1.5, group member 1 = 0.5, group member 2 = 0.5. Suppose that, you contribute 15 tokens to the group project, group member 1 contributes 5 tokens to the group project, and group member 2 contributes 10 tokens to the group project. The earnings in tokens of each of the participants are given by:

$$20 - \text{tokens contributed} + \text{multiplication factor} \times \text{sum of all contributions}$$

In your case this equals:  $20 - 15 + 1.5 \times (15 + 5 + 10) = 50$  tokens.

For group member 1 this equals:  $20 - 5 + 0.5 \times (15 + 5 + 10) = 30$  tokens.

For group member 2 this equals:  $20 - 10 + 0.5 \times (15 + 5 + 10) = 25$  tokens.

## The second stage

At the beginning of the second stage, everyone in the group will see how much each of the other group members contributed to the project as well as their earnings from the first stage. The decision each group member has to make in the second stage is to either reduce or leave equal the earnings of each other group member. Reducing other group members' earnings can be done by spending tokens. The other group members can also reduce your earnings if they wish to. All decisions are made simultaneously. That is, nobody will be informed about the decision of the other group members before everyone made his or her decision.

More concisely, in this stage, you must decide whether and if yes how many tokens you want to spend to reduce the earnings of the other two group members. If you want to reduce another member's earnings, you do that by allocating deduction points. For each deduction point that you allocate to another group member his or her earnings are reduced by 3 tokens and your own earnings are reduced by 1 token. If you do not wish to change the earnings of another group member then you must allocate 0 deduction points to him or her. Note, that you will not be allowed to reduce the earnings of a group member to less than zero.

Remember that, for every deduction point you receive from other group members, your earnings will be reduced by 3 tokens (but never below zero). Every participant can spend up to a maximum of 10 tokens (i.e. allocate 10 deduction points) on each group member in each period.

After everyone has made a decision, you will be informed how many deduction points you received from the other group members and also what your total earnings in tokens for that period are. Note that you do not get to know how individual group members spend their deduction points. In other words, you will only be informed of the total amount of deduction points allocated to you by the other two group members. You will not know how many deduction points each individual group member allocated to you.

## Examples for the second stage

Here are some arbitrarily chosen examples that illustrate how your final earnings are calculated. You, group member 1 and group member 2 are all members of the same group.

*Example 1:* Suppose that after the first stage you have earnings that are equal to 30 tokens. In the second stage you decide to allocate 3 deduction points to group member 1 (this reduces group member 1's earnings by 9 tokens) and 0 deduction points to group member 2 (this does not change group member 2's earnings). After all have made their decision, you learn that the others allocated you a total of 4 deduction points. In this case, your total earnings in tokens in this period are given by:

$$\begin{aligned} & (\text{Your first stage earnings} - 3 \times \text{deduction points allocated to you})^* \\ & \quad - \text{deduction points you allocated} \end{aligned}$$

\* If the number between brackets is negative then replace it with zero.

In this example, your earnings are equal to:  $(30 - 3 \times 4) - 3 = 18 - 3 = 15$  tokens.

*Example 2:* Suppose that after the first stage you have earnings that are equal to 18 tokens. In the second stage you decide to allocate 4 deduction points to group member 1 (this reduces group member 1's earnings by 12 tokens) and 6 deduction points to group member 2 (this reduces group member 2's earnings by 18 tokens). After all have made their decision, you learn that the others allocated you a total of 8 deduction points.

In this case, your earnings are equal to:  $(18 - 3 \times 8) - 10 = 0 - 10 = -10$  tokens.

Note that  $18 - 3 \times 8 = -6$ , since this is a negative number it is replaced by zero.

### **Negative earnings**

It is, in principle, possible that you make negative earnings in a period. However, you can always avoid this by not spending any tokens in the second stage (that is, by not allocating any deduction points to the other members). Hence, you can always avoid negative earnings with certainty through your own choices.

### **Summary**

In summary, your earnings in tokens in each period are equal to:

$$\begin{aligned} & (\text{Your initial endowment} - \text{your contribution to the project} \\ & \quad + 0.5 \times (\text{sum of contributions}) \\ & \quad - 3 \times \text{total deduction points received from others})^* \\ & \quad - \text{amount of deductions points you allocated to others} \end{aligned}$$

\* If your earnings up to this point are negative then replace them with zero

## B Descriptive Statistics

Table 3 summarizes the average amount contributed to the public good per period. Table 4 summarizes the average amount contributed to the public good per period by low-benefit and high-benefit subjects in privileged groups.

Table 3: Mean contributions by group type and treatment

Period \ Group	<i>Baseline Treatment</i>				<i>Punishment Treatment</i>			
	Normal		Privileged		Normal		Privileged	
1	10.67	(7.79)	11.96	(6.99)	12.83	(5.97)	10.44	(7.15)
2	9.05	(7.24)	11.67	(7.25)	13.94	(6.22)	12.89	(7.11)
3	5.14	(5.33)	9.25	(7.61)	15.94	(4.92)	10.67	(8.10)
4	3.10	(3.60)	7.92	(7.61)	15.39	(6.02)	10.89	(7.27)
5	3.00	(3.48)	8.46	(8.16)	17.28	(3.46)	13.17	(7.69)
6	2.52	(3.70)	8.46	(7.68)	17.50	(3.49)	14.06	(7.91)
7	2.71	(4.01)	6.88	(7.64)	15.83	(6.61)	14.11	(7.03)
8	2.67	(3.32)	8.25	(8.52)	16.83	(4.58)	11.94	(8.56)
9	2.24	(3.19)	7.13	(8.82)	16.83	(4.97)	12.28	(8.78)
10	1.00	(2.47)	6.79	(9.56)	14.89	(6.91)	13.11	(8.85)
Total	4.21	(5.52)	8.68	(8.05)	15.73	(5.49)	12.36	(7.78)

*Note:* Numbers between brackets are standard deviations.

Table 4: Mean contributions in privileged groups by subject type and treatment

Period \ Treatment	<i>Low-benefit Subjects</i>				<i>High-benefit Subjects</i>			
	Baseline		Punishment		Baseline		Punishment	
1	8.63	(5.80)	6.50	(4.83)	18.63	(3.5)	18.33	(3.20)
2	8.13	(5.88)	10.08	(6.80)	18.75	(3.54)	18.50	(3.67)
3	5.63	(4.72)	8.50	(7.35)	16.50	(7.23)	15.00	(8.37)
4	4.38	(4.32)	7.83	(6.41)	15.00	(8.02)	17.00	(4.69)
5	3.75	(4.71)	9.83	(7.42)	17.88	(4.36)	19.83	(0.41)
6	4.88	(4.95)	11.17	(8.33)	15.63	(7.29)	19.83	(0.41)
7	3.56	(3.74)	11.25	(7.03)	13.50	(9.30)	19.83	(0.41)
8	4.56	(6.03)	8.33	(8.29)	15.63	(8.21)	19.17	(2.04)
9	1.31	(2.33)	10.00	(8.64)	18.75	(3.54)	16.83	(7.76)
10	0.19	(0.75)	9.67	(9.07)	20.00	(-)	20.00	(-)
Total	4.50	(5.12)	9.32	(7.36)	17.02	(6.13)	18.43	(4.23)

*Note:* Numbers between brackets are standard deviations.

## C Regressions

This section contains the regressions referred to in Section 3.2 on which Figure 3 is based. Table 5 shows the results for normal groups and Table 6 for privileged groups. The regressions estimate the effect of being punished in period  $t - 1$  on the change in contribution from period  $t - 1$  to period  $t$ . We employ an OLS regressions with the change in contribution from period  $t - 1$  to period  $t$  as the dependent variable ( $\Delta c_i$ ). As independent variables we include a binary variable indicating whether a subject is punished or not, a dummy variable to control for a possible end game effect, and a series of variables measuring the subject's contribution relative to that of others in period  $t - 1$ . Specifically, in normal groups we use the (positive or negative) difference between  $i$ 's contribution and the group's median contribution. In privileged groups, we separate this variable into the difference between  $i$ 's contribution and the contribution of the high-benefit subject ( $c^H$ ) and the difference between  $i$ 's contribution and the contribution of the other low-benefit subject ( $c^L$ ). In addition, we interact these variables with the binary punishment variable. This model allows us to explore two effects. First, whether receiving punishment affects contributions and, secondly, whether the effect of punishment depends on the contribution of the punished subject relative to others in the group.<sup>16</sup> The results reported in Figure 3 consist of the linear combinations of coefficients that sums these two effects at various relative contribution levels. The concrete form of these linear combinations depends on the investigated situation. For example, for normal groups, if a subjected contributed 5 tokens less than the median contributor then the total effect is given by the sum of coefficients of the following variables: 'Received punishment' +  $5 \times$  'Received punishment  $\times$  negative deviation from  $c^{median}$ '. Similarly, for low-benefit subjects in privileged groups, if the subject in question contributed 5 tokens less than the other low-benefit subject the total effect is calculated as: 'Received punishment' +  $10 \times$  'Received punishment  $\times$  negative deviation from  $c^H$ ' +  $5 \times$  'Received punishment'  $\times$  negative deviation from  $c^L$ . We use robust standard errors and clustering on each group and period.

---

<sup>16</sup>If instead of a binary variable for punishment we run the same regressions with the number of punishment points, we get similar results but with a worse fit. In particular, the punishment variables lose significance. This suggests that the fact of being punished is more important than the amount of punishment one receives.

Table 5: Change in contribution depending on the previous period (Normal Groups)

Dependent variable: $\Delta c_i$	<i>Normal Groups</i>
Change in $i$ 's contribution	All
Received punishment	1.301
1 if $p_{ji} + p_{ki} > 0$	(0.838)
Positive deviation from $c^{median}$	-0.636***
$\max(c_i - c^{median}, 0)$	(0.199)
Received punishment $\times$ positive deviation from $c^{median}$	-0.593
$\max(c_i - c^{median}, 0) \times 1$ if $p_{ji} + p_{ki} > 0$	(0.618)
Negative deviation from $c^{median}$	-1.292
$\max(c^{median} - c_i, 0)$	(0.926)
Received punishment $\times$ negative deviation from $c^{median}$	1.786*
$\max(c^{median} - c_i, 0) \times 1$ if $p_{ji} + p_{ki} > 0$	(0.939)
Last period	-1.873
1 if $t = 10$	(1.386)
Constant	-0.312
	0.312
# Observations	162
$R^2$	0.318

*Note:* Numbers between brackets are standard errors \*\*\* Significant at the 1% level.

\*\* Significant at the 5% level. \* Significant at the 10% level.

Table 6: Change in contribution depending on the previous period (Low-benefit subjects)

Dependent variable: $\Delta c_i$	<i>Privileged Groups</i>
Change in $i$ 's contribution	Low-benefit
Received punishment	4.015**
1 if $p_{ji} + p_{ki} > 0$	(1.704)
Positive deviation from $c^H$	-0.100
$\max(c_i - c^H, 0)$	(0.356)
Received punishment $\times$ positive deviation from $c^H$	-0.131
$\max(c_i - c^H, 0) \times 1$ if $p_{ji} + p_{ki} > 0$	(0.618)
Negative deviation from $c^H$	0.085
$\max(c^H - c_i, 0)$	(0.066)
Received punishment $\times$ negative deviation from $c^H$	-0.091
$\max(c^H - c_i, 0) \times 1$ if $p_{ji} + p_{ki} > 0$	(0.108)
Positive deviation from $c^L$	-0.153
$\max(c_i - c^L, 0)$	(0.142)
Received punishment $\times$ positive deviation from $c^L$	-0.409
$\max(c_i - c^L, 0) \times 1$ if $p_{ji} + p_{ki} > 0$	(0.511)
Negative deviation from $c^L$	0.198
$\max(c^L - c_i, 0)$	(0.148)
Received punishment $\times$ negative deviation from $c^L$	0.112
$\max(c^L - c_i, 0) \times 1$ if $p_{ji} + p_{ki} > 0$	(0.218)
Last period	-0.116
1 if $t = 10$	(1.166)
Constant	-1.900
	1.400
# Observations	108
$R^2$	0.286

Note: Numbers between brackets are standard errors \*\*\* Significant at the 1% level.

\*\* Significant at the 5% level. \* Significant at the 10% level.



## References

- Babcock, L., Camerer, C. F., and Samuel, S. I. (1995). Biased judgments of fairness in bargaining. *American Economic Review*, 85:1337–1343.
- Benkler, Y. (2002). Coase’s penguin, or, linux and *The Nature of the Firm*. *The Yale Law Journal*, 112:369–446.
- Bochet, O., Page, T., and Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization*, 60:11–26.
- Bolton, G. E. and Ockenfels, A. (2000). A theory of equity, reciprocity, and competition. *American Economic Review*, 90:166–193.
- Bowles, S. and Gintis, H. (2005). Pro-social emotions. In Blume, L. E. and Durlauf, S. N., editors, *The Economy as a Complex Evolving System III: Essays in Honor of Kenneth Arrow*. Oxford University Press, Oxford.
- Brandts, J. and Schram, A. (2001). Cooperation and noise in public goods experiments: applying the contribution function approach. *Journal of Public Economics*, 79:399–427.
- Carpenter, J. P. (2006). The demand for punishment. *Journal of Economic Behavior & Organization*, 62:522–542.
- Charness, G. and Levine, D. K. (2006). Intention and stochastic outcomes: An experimental study. *The Economic Journal*, forthcoming.
- Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117:817–869.
- Cinyabuguma, M., Page, T., and Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics*, 9:265–279.
- Egas, M. and Riedl, A. (2005). The economics of altruistic punishment and the demise of cooperation. Discussion Papers 05-065/1, Tinbergen Institute.
- Engelmann, D. and Strobel, M. (2004). Inequality aversion, efficiency and maximim preferences in simple distribution experiments. *American Economic Review*, 94:857–869.
- Falk, A., Fehr, E., and Fischbacher, U. (2000). Testing theories of fairness: Intentions matter. Working Paper 63, University of Zurich.
- Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54:293–315.
- Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90:980–994.
- Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415:137–140.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114:817–868.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10:171–178.
- Fisher, J., Isaac, M. R., Schatzberg, J. W., and Walker, J. M. (1995). Heterogeneous demand for public goods: Effects on the voluntary contributions mechanism. *Public Choice*, 85:249–266.

- Gächter, S. and Herrmann, B. (2006). The limits of self-governance in the presence of spite: Experimental evidence from urban and rural russia. Discussion paper no. 2006-13, CeDEx.
- Gächter, S. and Riedl, A. (2005). Moral property rights in bargaining with infeasible claims. *Management Science*, 51(2):249–263.
- Hopfensitz, A. and Reuben, E. (2005). The importance of emotions for the effectiveness of social punishment. Discussion paper 05-075, Tinbergen Institute.
- Isaac, M. R., Walker, J. M., and Thomas, S. (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice*, 43:113–149.
- Johnson, R. N. and Libecap, G. D. (1982). Contracting problems and regulation: The case of fishery. *American Economic Review*, 72(5):1005–1022.
- Knez, M. J. and Camerer, C. F. (1995). Outside options and social comparison in a three-player ultimatum game experiments. *Games and Economic Behavior*, 10:65–94.
- Ledyard, J. O. (1995). Public goods: A survey of experimental research. In Kagel, J. H. and Roth, A. E., editors, *The Handbook of Experimental Economics*, pages 111–194. Princeton University Press, New Jersey.
- Levin, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1:593–622.
- Masclot, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). Monetary and non-monetary punishment in the voluntary contribution mechanism. *American Economic Review*, 93:366–380.
- McCabe, K. A., Rigdon, M. L., and Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, 52:267–275.
- Olson, M. (1980[1965]). *The logic of collective action*. Harvard University Press, Cambridge.
- Ostrom, E., Walker, J. M., and Gardner, R. (1992). Covenants with and without a sword: Self governance is possible. *American Political Science Review*, 86:404–417.
- Palfrey, T. R. and Prisbrey, J. E. (1996). Altruism, reputation and noise in linear public good experiments. *Journal of Public Economics*, 61:409–427.
- Palfrey, T. R. and Prisbrey, J. E. (1997). Anomalous behavior in public goods experiments: How much and why? *American Economic Review*, 87:829–846.
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press, Cambridge.
- Reuben, E. and van Winden, F. (2005). Negative reciprocity and the interaction of emotions and fairness norms. Discussion paper 05-014, Tinbergen Institute.
- Saijo, T. and Nakamura, H. (1995). The spite dilemma in voluntary contribution mechanism experiments. *Journal of Conflict Resolution*, 39:535–560.
- Sandler, T. (1992). *Collective Action: Theory and Application*. The University of Michigan Press.
- Siegel, S. and Castellan, N. (1988). *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, New York.
- Zelmer, J. (2003). Linear public goods experiments: A meta-ananlysis. *Experimental Economics*, 6:299–310.