Endogenous Institution Choice in Social Dilemmas

Özgür Gürerk

2007

DISSERTATION

zur Erlangung des Grads eines Doktors der Wirtschaftswissenschaft
(Dr. rer. pol.) der

Universität Erfurt

Staatswissenschaftliche Fakultät

Gutachter: Prof. Dr. Bettina Rockenbach

Prof. Dr. Manfred Königstein

Datum der Disputation: 14.11.2007

urn:nbn:de:gbv:547-200701153
[http://nbn-resolving.de/200701153]

Zusammenfassung

Die ökonomische Theorie kann viele Facetten des kooperativen Verhaltens, die durch einen Konflikt zwischen Individual- und Gemeinschaftsinteressen gekennzeichnet sind, wie z.B. Kooperation in sozialen Dilemmasituationen, nicht umfassend erklären. Empirische Evidenz aus experimentellen Untersuchungen zeigt, dass sogar elementare institutionelle Rahmenbedingungen wie einfache Belohnungs- oder Bestrafungsmechanismen Kooperation fördern können, falls diese in einer Gesellschaft von Außen (exogen) „installiert" werden. Die Frage, ob diese Sanktionsmechanismen von Individuen selbst (endogen) gewählt und einen Wettbewerb mit alternativen Institutionen „überleben" würden, ist noch nicht ausreichend beantwortet worden. Diese Dissertation trägt zur Kooperationsforschung bei, indem sie die endogene Wahl von Institutionen wie Bestrafungs- und Belohnungsmechanismen und deren Auswirkungen auf Kooperationsverhalten in sozialen Dilemmasituationen untersucht.

Schlagworte: endogen, Institution, Wahl, soziales Dilemma, Experiment, Bestrafung, Belohnung

Abstract

Standard economic theory fails to explain many facets of cooperative behavior, e.g., cooperation in social dilemma situations which are characterized by the conflict between individual and collective interests. There is empirical evidence from experimental studies showing that even elementary institutional arrangements such as simple reward or punishment mechanisms can foster cooperation when these mechanisms are exogenously "installed" in a community. The question of whether these sanctioning mechanisms would be endogenously chosen by individuals and "survive" a competition with alternative institutions is not yet satisfactorily answered. This thesis contributes to the literature on cooperation by investigating the endogenous choice of institutions with punishment and reward mechanisms, and by exploring their effects on cooperative behavior in social dilemma situations.

Keywords: endogenous, institution, choice, social dilemma, experiment, punishment, rewards

To my beloved mother Serap Gürerk

Acknowledgements

First of all, I am greatly indebted to my advisor, Bettina Rockenbach, for her continuous support and for the provision of an excellent working atmosphere.

I would like also thank my co-author Bernd Irlenbusch for many inspiring discussions.

I also thank my current and former colleagues at the University of Erfurt; Kai Ahlborn, Pierre Gericke, Stefan Große, Mario Gruppe, Sebastian Händschke, Mareike Hoffmann, Vahidin Jeleskovic, Manfred Königstein, Thomas Lauer, Mark Meyer, Mark Peacock, Christiane Pilz, Elke Renner, Arne Weiß, Tim Wenniges, and Irenaeus Wolff, for their valuable comments and help.

My very special thanks go to Andrea Bäcker.

Most of all I am very grateful for the everlasting support, encouragement, and love that I received from my parents, Serap Gürerk and Alim Gürerk, and my brother Barış Can Gürerk.

CONTENTS

Overview

The phenomenon of human cooperation is object of the study in many research disciplines. While evolutionary models based on kinship (Hamilton, 1964) and direct reciprocity (Trivers, 1971) explain cooperation among relatives and in direct relationships, the emergence and evolution of cooperation in large groups is still not fully understood. Standard economic theory fails in explaining many facets of cooperative behavior, e.g., cooperation in social dilemma situations which are characterized by the conflict between individual and collective interests. In contrast to theoretical predictions, experimental evidence shows the existence of cooperation even in social dilemma situations (see e.g. Ledyard, 1995 or Camerer, 2003). However, if institutional rules are absent, cooperation is highly fragile.

Institutions can be broadly defined as "the prescriptions that humans use to organize all forms of repetitive and structured interactions" (Ostrom, 2005). The institutional framework defines "the rules of the game" on which the monetary and non-monetary consequences of individuals' actions depend. It is obvious that institutional arrangements influence the behavior of the interacting individuals in social dilemma situations. For this reason it is of great interest to study the relationship between the institutional framework and human behavior in social dilemmas. If we can understand *how institutions work* and *how they influence the human behavior* we will be able to design efficient institutions that minimize the downside effects that are inherent to social dilemmas.

There exists theoretical and empirical evidence that even elementary institutional arrangements such as simple reward or punishment mechanisms can foster cooperation once these mechanisms are exogenously "installed" in a community (see e.g., Yamagishi, 1986, Ostrom et al., 1992, Fehr and Gächter, 2000, Sefton et al., forthcoming). In these studies, however, members neither have the possibility to abstain from the interaction nor the opportunity to leave the community in order to establish a new community with different institutional arrangements. The question of whether sanctioning mechanisms would be endogenously chosen and "survive" a competition with alternative institutions is not yet satisfactorily answered.

This thesis contributes to the cooperation research by investigating the *endogenous choice* of institutions and their effects on cooperative behavior in social dilemma situations. The competition between institutions is modeled by letting individuals *endogenously* decide under which institutional arrangements they want to interact with other individuals in a social dilemma situation. In other words, individuals have the opportunity to *vote with their feet* by voluntarily joining the preferred institution or community.

Chapter 1 investigates experimentally whether a community with sanctioning possibilities performs better than a community without any sanctioning mechanism. Subjects are repeatedly given the choice between a community with reward *and* punishment possibilities and a community without any sanctioning options, i.e., a pure voluntary contribution mechanism. Initially, most subjects are reluctant to join the community with sanctioning possibilities. However, individuals with a strong disposition to cooperate manage to establish a "cooperative culture" by contributing high to the joint project and punishing the free-riders harshly. The heavy punishment in the beginning causes efficiency losses in the short run, but ensures that the sanctioning community becomes more profitable than the non-sanctioning community in the long-run. Hence, the sanctioning community attracts more and more subjects. In the end, the overwhelming majority of subjects inhabit the sanctioning community while the non-sanctioning community gets depopulated. Moreover, the

cooperation flourishes in the sanctioning community. Virtually all subjects contribute their whole endowment to the public good. A credible punishment threat makes actual punishment unnecessary. Thus the ideal of the social optimum is reached.

Chapter 2 deepens the analysis of the sanctioning mechanisms and their effects on cooperation theoretically and experimentally. We disentangle the effects of rewards and punishment: in an experiment subjects are given the choice to choose between the non-sanctioning community and a community with a unique sanction option. In one treatment, a rewarding community is the alternative to the non-sanctioning community; in the other treatment, a punishment community is the alternative. All three institutional arrangements, i.e., the non-sanctioning community, and the communities with reward and punishment possibilities, are investigated theoretically. To analyze the community choice we develop a simple concept of community choice equilibrium based on the inequality aversion theory by Fehr and Schmidt (1999). The experimental results are clear cut. While the punishment community performs as well as the community with the combined sanctioning possibilities from Chapter 1, the reward community performs poorly. Although, there is a relatively high cooperation level in the reward community initially, over time it deteriorates to rather low levels. However, as in case of the institution with combined sanctioning possibilities from the first chapter, the pure punishment community also suffers from considerable initial efficiency losses due to heavy punishment. How can these initial efficiency losses be mitigated?

Chapter 3 deals with this question. One possible explanation for the initial reluctance and losses could be that subjects simply not anticipate correctly that the punishment community is the more efficient community in the long run. Previous studies show that provision of a social history about the results of an experiment may indeed influence the behavior of the informed subjects who are going to play the same game (Berg et al., 1995). Thus, in Chapter 3, we replicate the treatment with the pure punishment community from Chapter 2 with the addition that subjects are given a social history. We conjecture that more information provided to participants referring to the superior performance of the punishment community may lower the reluctance against it, thus lowering the initial efficiency losses. Indeed, we find that informed subjects are significantly less reluctant to join the punishment community. Moreover, the social optimum of full cooperation is established more rapidly and there are less efficiency losses due to reduced punishment activity.

The findings from Chapter 1-3 indicate that there exists a fundamental difference between the use of rewards and punishment. While punishment is an effective instrument exerted by contributors to discipline free-riders, rewards are usually given by cooperators to those who already cooperate. Thus, one possible reason for the poor performance of the reward community could be that only the “good” reward each other. The question arises whether the reward mechanism would perform better in a hierarchical framework, e.g., if a senior member of the community or a leader in a work team possesses the discriminatory power to reward or to punish.

In Chapter 4, we experimentally investigate in a team-work setting which mechanism team leaders prefer when they have the choice to apply a positive incentive scheme (rewards) or a negative incentive scheme (punishment), and how different incentive schemes influence the performance of teams. Dependent on the chosen incentive scheme, the leader can individually reward or punish his teammates. Initially, almost all leaders prefer using positive incentives. However, this preference changes over time and finally the majority of the leaders opt for negative incentives. Initially, teams’ overall performances are higher under the positive incentive scheme than under the negative incentive scheme. However, after a while, the teams under the negative incentive scheme perform better. In the end, cooperation under the positive

incentive scheme breaks down completely whereas in teams experiencing the negative incentive scheme leaders as well as teammates maintain high contribution levels.

References

Camerer, Colin. “Behavioral Game Theory: Experiments on Strategic Interaction.” Princeton: Princeton University Press, 2003.

Fehr, Ernst and Gächter, Simon. “Cooperation and Punishment in Public Goods Experiments.” *American Economic Review*, 2000, 90(4), pp. 980-994.

Hamilton, W.D. “The genetical evolution of social behaviour. I.” *Journal of Theoretical Biology*, Volume 7, Issue 1, July 1964, Pages 1-16.

Ledyard, J. “Public Goods: A Survey of Experimental Research.” J. Kagel and A. Roth, Handbook of Experimental Economics, 1995, Princeton University Press, pp. 111-194.

Ostrom, E.; Walker, J. and Gardner, R. "Covenants with and without a Sword - Self-Governance Is Possible." *American Political Science Review*, 1992, *86*(2), pp. 404-17.

Ostrom, E. “Understanding Institutional Diversity.” Princeton: Princeton University Press, 2005.

Sefton, M.; Shupp, R., and Walker, J. “The Effect of Rewards and Sanctions in the Provision of Public Goods.” forthcoming in *Economic Inquiry*.

Trivers, R. L. “Evolution of reciprocal altruism.” *Quarterly Review of Biology*, 1971 Vol. 46 (1), pp. 35-57.

Yamagishi, T. “The Provision of a Sanctioning System as a Public Good.” *Journal of Personality and Social Psychology*, 1986, *51*(1), pp. 110-16.

1. The competitive advantage of sanctioning institutions[1]

1.1. Introduction

Understanding the fundamental patterns and determinants of human cooperation and the maintenance of social order in human societies is a challenge across disciplines. The existing empirical evidence for the higher levels of cooperation when altruistic punishment is present versus when it is absent systematically ignores the institutional competition inherent in human societies. Whether punishment would be deliberately adopted and would similarly enhance cooperation when directly competing with non-punishment institutions is highly controversial in light of recent findings on the detrimental effects of punishment. We show experimentally that a sanctioning institution is the undisputed winner in a competition with a sanction-free institution. Despite initial aversion, the entire population migrates successively to the sanctioning institution and strongly cooperates while the sanction-free society becomes fully depopulated. The findings demonstrate the competitive advantage of sanctioning institutions and exemplify the emergence and manifestation of social order driven by institutional selection.

The uniqueness of human cooperation necessitates investigations that reach beyond the explanations of cooperative behavior of non-human animals (see e.g. Stevens and Hauser, 2004, Fehr and Fischbacher, 2003, Henrich et al., 2001, Ostrom et al., 1999, Hammerstein, 2003). Profound empirical evidence shows that the possibility of sanctioning norm violators stabilizes human cooperation at a high level whereas cooperation typically collapses in the absence of sanctioning possibilities (see e.g. Yamagishi, 1986, Fehr and Gächter, 2002, Ostrom et al., 1992, Andreoni et al., 2003, Masclet et al., 2003, Rege and Telle, 2004). Would a sanctioning institution deliberately be adopted when individuals can choose between a sanctioning and a sanction-free institution? The considerable payoff losses in the process towards stable cooperation – for both the punishers and the punished individuals – as well as natural resentments against punishment caused for example by its detrimental effects might guide individuals' choice towards the sanction-free institution (Fehr and Rockenbach, 2003).

The argument that higher cooperation levels in sanctioning institutions "automatically" lead to their prevalence – because rational individuals choose the institution with the higher payoff (Binmore, 2005) – is often brought forward as an affirmative argument for the competitive advantage of sanctioning institutions. The force of this argument can be questioned, however, since it displaces rather than solves the evolutionary puzzle of human cooperation. The reason for this is that stable cooperation requires a positive share of individuals who carry personal costs for cooperation and punishment to the benefit of the entire group (Gintis, 2000, Fehr et al., 2002, Gintis et al., 2003). These individuals have a clear payoff disadvantage compared to cooperators who free-ride on the punishment acts. Recent research shows that a positive share of strong reciprocators – cooperating individuals who are willing to reward fair behavior and to punish unfair behavior even when they cannot gain materially from doing so – can be evolutionarily stable (Boyd et al., 2003, Bowles and Gintis, 2004). But what happens if the population is perfectly mobile and is permanently invaded by outsiders from a non-cooperative environment who are attracted by high payoffs from cooperation? Is the fraction of strong reciprocators who choose the sanctioning institution sufficiently large to keep up the cooperative culture? These arguments cast serious doubt on the prevalence of sanctioning institutions.

[1] This chapter is based on the article "The Competitive Advantage of Sanctioning Institutions" published in *Science*, 2006, Vol. 312(5770), pp. 108-111, joint work with Bernd Irlenbusch and Bettina Rockenbach. All authors contributed equally.

However, several affirmative arguments for the competitive advantage of sanctioning institutions also come to mind, e.g. the considerable frequency of institutional frameworks which facilitate the sanctioning of norm violators in human societies (Mahdi, 1986, Johnson and Earle, 1987, Wiessner, 2005) and the recent finding that humans derive satisfaction from punishing defectors (de Quervain et al., 2004). Additionally, theories of cultural and institutional selection (Boyd and Richerson, 1992, Henrich and Boyd, 2001, Boyd and Richerson, 2002, Henrich, 2004) that ground on the exceptional human ability of social learning support the competitive advantage of sanctioning institutions. They suggest that individuals preferentially migrate to groups with higher payoffs and imitate the decisions prevalent in these groups. Hence, group members punish, because it is common to do so. When cooperation is sufficiently widespread, the payoff-disadvantage from punishing is relatively small and only a weak tendency for conformist behavior suffices to stabilize the punishment of non-cooperators.

We inquire into the competitive advantage of sanctioning institutions in a laboratory experiment with the novel feature of implementing permanent competition between a sanctioning and a sanction-free institution through endogenous choice. It allows one to study the evolution of the different institutions over time as well as the changes in behavior in the same individual when participating in different social settings.

1.2. Experimental Design

In our experiment, 84 participants anonymously interact in a social dilemma situation in 30 repetitions. Each repetition consists of three stages: An institution choice stage (S0), a voluntary contribution stage (S1), and a sanctioning stage (S2). In stage S0, the participants simultaneously and independently choose between a sanctioning institution (SI) and a sanction-free institution (SFI) in which neither positive sanctioning (rewards) nor negative sanctioning (punishment) is possible. In stage S1, each participant interacts in a public goods game with all other participants who have chosen the same institution in S0: each player is endowed with 20 money units (MUs) and may contribute between 0 and 20 MUs to a public good. Each group member equally profits from the public good, independent from his or her own contribution. The MUs not contributed to the public good are transferred to the participant's private account. The diametrically opposed individual and collective interests constitute the social dilemma in public good provision: It is always in the material self-interest of any subject to free-ride on the contributions of others and to keep all MUs for the private account, while the collective interest demands full contribution of all group members. After the players have simultaneously made their contribution decisions, they are informed about the contributions of each member in the own group. In stage S2 each player in SI may positively or negatively sanction other members of SI by assigning between zero and 20 tokens to other members. Each token employed as a negative sanction costs the punished member 3 MUs and the punishing member 1 MU. Each token employed as a positive sanction yields the receiving member 1 MU and costs the employing member 1 MU. At the end of the period each participant receives detailed (but anonymous) information about each of the other participants from both institutions (see the Appendix for more details on the methods).

1.3. Results

The initial choice of institution provides a clear picture: only about one third of the participants (mean 36.9 percent; standard error 4.0 percent) prefer SI to SFI in the first period. The revealed institution preference correlates with different types of behavior (see also Fischbacher et al., 2001, Kurzban and Houser, 2005). Participants who initially join SI contribute on average 12.7 MUs (standard error 0.79) in the first period, while on average only 7.3 MUs (standard error 0.54) are contributed in SFI (Wilcoxon signed rank matched

pairs test, z = −2.366, P = 0.016, two-tailed). Almost half the subjects (mean 48.4 percent; standard error 8.5 percent) who opt for SI in the first period are "high contributors" in the sense that they contribute at least 15 MUs. Almost three fourths (mean 73.3 percent; standard error 17.0 percent) of these high contributors exert punishment tokens in order to discipline low contributors and thus try to enforce and establish a norm of high cooperation. These subjects amount to 13.1 percent (standard error 4.0 percent) of the total subject population and can clearly be classified as "strong reciprocators", i.e. subjects with a predisposition to make high contributions and to punish norm-violators. In contrast, 16.1 percent (standard error 5.2 percent) of the subjects in SI contribute 5 MUs or less ("free-riders") in the first period.

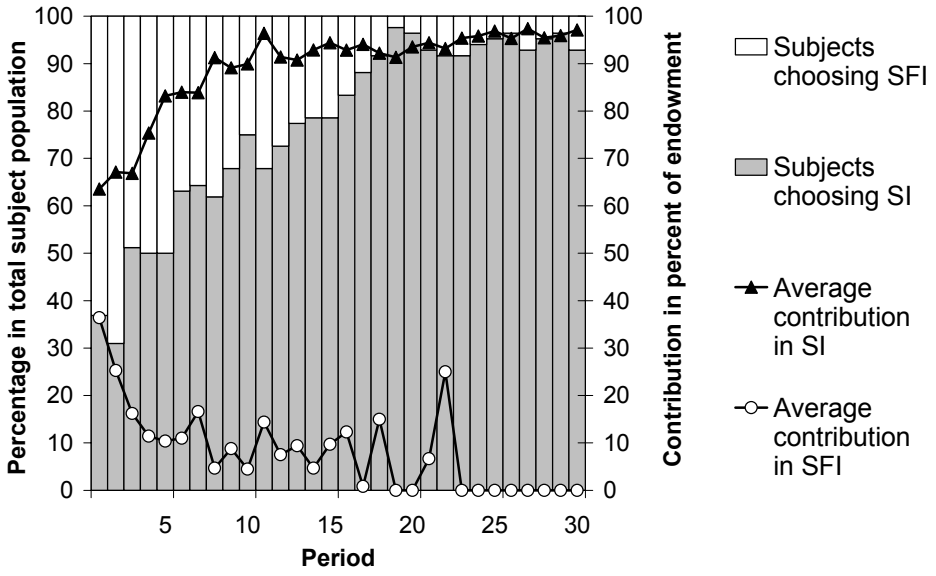


Figure 1.1 Subjects' choice of institution and their contributions. Figure 1.1 displays the average contributions in both institutions over the 30 periods of the interaction measured as the percentage of endowment contributed to the public good.

The situation is completely different in SFI, where in the first period almost half of the subjects are free-riders (mean 43.4 percent; standard error 3.4 percent) while high contributors are rare (mean 11.3 percent; standard error 4.3 percent). A subject who chooses SFI in the first period with a contribution of more than 15 MUs and employs negative sanctions immediately after having switched to SI may also be classified as a strong reciprocator. We observe exactly two subjects with this behavior in our subject population (2.4 percent), so that 15.5 percent (standard error 5.6 percent) is a lower bound for the proportion of strong reciprocators in the subject population. Initially, the significantly higher contributions in SI do not result in higher payoffs in SI: Average payoffs in the first period of SI (mean 38.1; standard error 2.05) are significantly lower than in SFI (mean 44.4; standard error 0.32) (Wilcoxon signed rank matched pairs test, z = −2.047, P = 0.047, two-tailed). Due to immense punishment activities, free-riders earn significantly less in SI (mean 30.2; standard error 4.51) than in SFI (mean 49.7; standard error 0.86) in the first period (Wilcoxon signed rank matched pairs test, z = −2.366, P = 0.016, two-tailed).

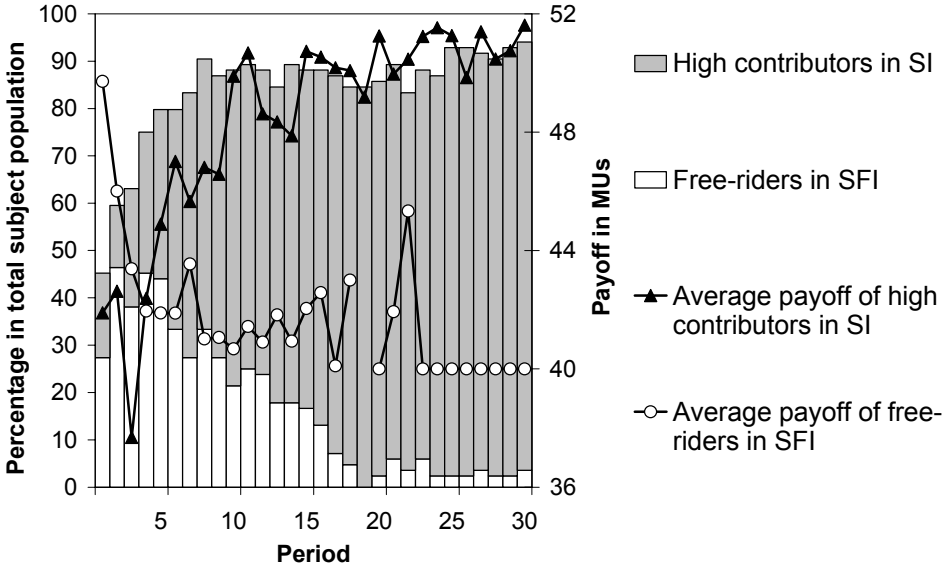


Figure 1.2 Payoffs of the two predominant behavioral patterns, "free-riders" (contributions between 0 and 5 MUs) in the sanction-free institution (SFI) and "high contributors" (contributions between 15 and 20 MUs) in the sanctioning institution (SI). Note that the highest attainable payoff (under full contributions of all subjects and no punishment) is 52 and the payoff from complete free-riding and no punishment is 40.

Although subjects are initially reluctant to join SI, it becomes predominant over time; in the end virtually all participants (mean 92.9 percent; standard error 3.4 percent) choose SI and cooperate fully (Figure 1.1).[2] Simultaneously, contributions in SFI decrease to a level of zero. In period 10 the contributions in SI are on average 89.9 percent (standard error 10.3 percent) of the endowment and from there on they steadily increase. In the last period the difference between the two institutions is almost as extreme as it can be with average contributions of 19.4 MUs (standard error 0.714) in SI and 0 MUs (standard error 0) in SFI. Averaged over all periods, subjects in SI contribute 18.3 MUs (91.4 percent of the endowment; standard error 5.0 percent), while subjects in SFI contribute only 2.9 MUs (14.4 percent of the endowment; standard error 3.0 percent) (Wilcoxon signed rank matched pairs test, $z = -2.366$, $P = 0.016$, two-tailed).

What causes this dramatic change of mind? Pure imitation of the successful behavior would lead to an increase of free-riders in SFI because they earn the highest average payoffs in the first period. This is actually observed in period two. As a consequence the payoffs of free-riders in SFI decrease and over the periods, participants in SFI experience the typically observed collapse of cooperation in repeated social dilemma interactions (Figure 1.1). A comparison of the payoffs of the two predominant behavioral patterns – free-riding in SFI and high contributions in SI (Figure 1.2) – shows that from period five onwards a high contributor in SI achieves a higher payoff than a free-rider in SFI (Wilcoxon signed rank matched pairs test, $z = -2.366$, $P = 0.016$, two-tailed). It therefore pays for a monetary payoff maximizing participant to switch from free-riding in SFI to contributing in SI. This triggers an amplifying effect, namely the greater the number of cooperators in SI, the higher their payoffs. Indeed from period 10 on, 86.1 percent (standard error 13.1 percent) of all members of SI contribute fully (20 MUs) and 86.0 percent (standard error 8.6 percent) in SFI contribute almost nothing (2 MUs or less). The finding that players apparently choose institutions according to payoffs indicates that stochastic forces play only a minor role in determining switching behavior (Young, 1998).

[2] Figure A1.1 in the supporting online material displays the exact flow in both directions between institutions from one period to the next.

A closer look at individual behavior immediately before and after migration from one institution to the other confirms the bipolar pattern of behavior induced by the two institutions. In fact, 80.3 percent (standard error 5.0 percent) of subjects increase their contribution when migrating from SFI to SI in two consecutive periods. And 27.1 percent (standard error 5.3 percent) of subjects even "convert" from being a complete free-rider (contributing 0 MUs) to a full cooperator (contributing 20 MUs) when switching from SFI to SI. The migration behavior in the opposite direction, i.e. from SI to SFI, is similarly extreme. Roughly 70 percent (mean 70.9 percent; standard error 4.9 percent) of subjects reduce their contribution when switching from SI to SFI and about 20 percent (mean 17.0 percent; standard error 4.7 percent) switch from full cooperation to free-riding.

Individual payoff maximization cannot explain why new members in SI follow the second norm established by the strong reciprocators who joined SI in early periods, i.e. the norm to punish low contributors. The most successful behavior would be to contribute in SI (and hence avoid being punished), but refrain from the costly punishment of others. Since punishment of defectors constitutes a second-order public good (in which defection cannot be sanctioned in our setting), individual payoff maximization would rule out punishment. However, only a minority of subjects follow this payoff maximizing behavior. The overwhelming majority of 62.9 percent (standard error 8.5 percent) of the subjects immediately conforms to and adopts the prevailing norm of punishment in SI, i.e. they always use punishment immediately after they switch to SI. This results in a quite stable proportion of roughly 40 percent (mean 42.1 percent; standard error 5.9 percent) of subjects who both contribute highly and punish during the last 20 periods (Figure 1.3).

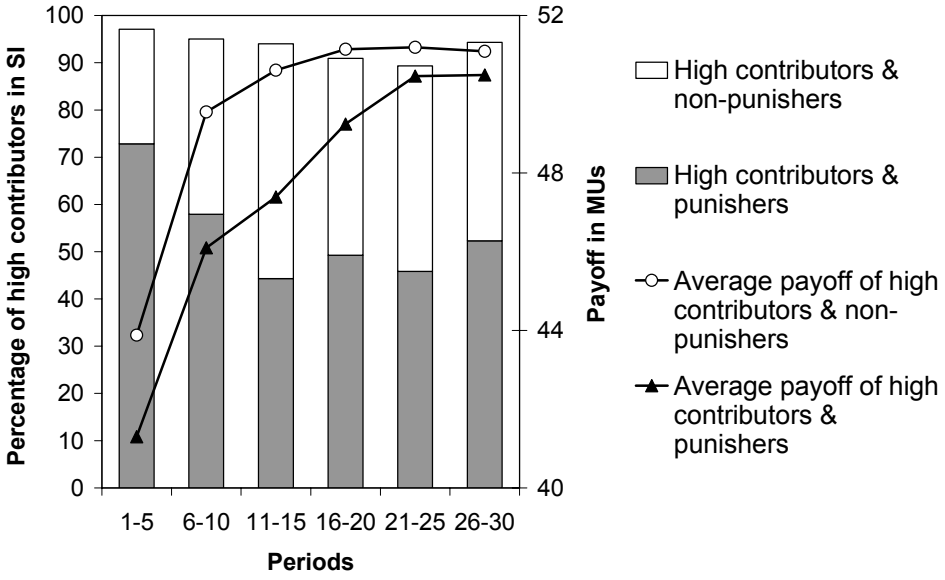


Figure 1.3 Payoffs and percentages of punishers and non-punishers among the "high contributors" (contributions between 15 and 20 MUs) in the sanctioning institution (SI). Note that the highest attainable payoff (under full contributions of all subjects and no punishment) is 52 and the payoff from complete free-riding and no punishment is 40.

Figure 1.3 also illustrates that the payoff difference between high contributors who punish and those who do not constantly diminishes over time because punishment becomes ever more unnecessary. Additionally, since the absolute number of punishers increases, the individual burden from effectively punishing free-riders becomes smaller over time[3]. Towards the end,

[3] A logistic regression shows that the stay duration in SI in terms of the number of periods has a significantly negative influence on the likelihood to punish others (see Table A1.1 in the Appendix). Note, however, that

subjects who both contribute highly and punish exhibit a payoff disadvantage of less than two percent, hence the "selection pressure" against strong reciprocators becomes quite weak.[4] This leads to a continuous increase in efficiency gains in SI up to 95.8 percent (standard error 4.6 percent) in the final period, whereas efficiency gains in SFI converge to zero (mean 0; standard error 0.0).

Although the use of both positive and negative sanctions per individual decreases over time, the ratio in which they are used is rather stable: on average 1.66 negative sanction points (standard error 0.60) are allocated per positive sanction point. A Tobit regression of the combined effect of positive and negative sanctions exhibits a clear positive impact of punishment on subsequent contributions, while positive sanctions have a slightly negative but rather insignificant effect (Table 1.1).

Table 1.1 Results of a Tobit regression, explained variable: Contribution (t+1) – Contribution (t)

	Coefficient	Z value
Negative sanctions in t	.444	5.24***
	(.085)	
Positive sanctions in t	-.148	-1.45
	(.102)	
Constant	.000	0.00
	(.053)	

Tobit regression for subjects who opted for SI in period t and t+1 with a robust estimation for the standard errors using the independent observations as clusters.
*** denotes significance at 1%. The values in parentheses denote the robust standard errors.

It seems that positive sanctions are not perceived as an unambiguous encouragement to increase the contribution; perhaps they are taken as an indication that the contribution has been higher than expected by others and hence may be lowered. These observations reflect the asymmetry between negative and positive sanctions. Positive sanctions are addressed to those who already abide by the social norm and in order to preserve the approval of cooperation, a continuous application of the instrument is required. Negative sanctioning, by contrast, is an instrument for disapproving of norm violating behavior and needs only be exerted if the norm is not followed. If an individual abides by the norm, punishment is not necessary. The threat of punishment alone is able to support cooperation.

1.4. Conclusions

Our results show that the sanctioning institution is the undisputed winner in a "voting-with-one's-feet" competition with a sanction-free institution. The results provide profound empirical evidence for the existence and importance of strong reciprocators as well as a form of conformist behavior as described in models of cultural selection. The initial establishment of the "norm to cooperate and punish free-riders" is mainly driven by the steadfastness of the strong reciprocators to punish non-cooperative subjects, despite severe individual losses (Camerer and Fehr, 2006). Although they are a minority, they manage to establish and enforce a cooperative culture which attracts even previously non-cooperative individuals and thus resolves the social dilemma. The predominant tendency to punish norm violators after a migration from the non-cooperative environment of the sanctioning-free institution to the sanctioning institution provides support for the assumption that humans adapt to the common

individually exerted punishment may be lowered over time to effectively punish a free-rider because the number of potential punishers becomes larger. In fact, average payoffs of free-riders decrease over periods as can be seen from Figure A1.2 in the Appendix.

[4] In the last ten periods subjects who contribute highly and punish reach on average 98.7 percent of the payoff of subjects who contribute highly but do not punish.

behavior although it deviates from the payoff maximizing behavior. This tendency for conformism raises sanctioning activities at a high level such that cooperation can be stabilized.

1.5. References

Andreoni, J.; Harbaugh, W. and Vesterlund, L. "The Carrot or the Stick: Rewards, Punishments, and Cooperation." *American Economic Review*, 2003, *93*(3), pp. 893-902.

Binmore, K. G. *Natural Justice*. New York: Oxford University Press, 2005.

Bowles, S. and Gintis, H. "The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations." *Theoretical Population Biology*, 2004, *65*(1), pp. 17-28.

Boyd, R.; Gintis, H.; Bowles, S. and Richerson, P. J. "The Evolution of Altruistic Punishment." *Proceedings of the National Academy of Sciences of the United States of America*, 2003, *100*(6), pp. 3531-35.

Boyd, R. and Richerson, P. J. "Group Beneficial Norms Can Spread Rapidly in a Structured Population." *Journal of Theoretical Biology*, 2002, *215*(3), pp. 287-96.

____. "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups." *Ethology and Sociobiology*, 1992, *13*(3), pp. 171-95.

Camerer, C. F. and Fehr, E. "When Does "Economic Man" Dominate Social Behavior?" *Science*, 2006, *311*(5757), pp. 47-52.

de Quervain, D. J. F.; Fischbacher, U.; Treyer, V.; Schelthammer, M.; Schnyder, U.; Buck, A. and Fehr, E. "The Neural Basis of Altruistic Punishment." *Science*, 2004, *305*(5688), pp. 1254-58.

Fehr, E. and Fischbacher, U. "The Nature of Human Altruism." *Nature*, 2003, *425*(6960), pp. 785-91.

Fehr, E.; Fischbacher, U. and Gachter, S. "Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms." *Human Nature-an Interdisciplinary Biosocial Perspective*, 2002, *13*(1), pp. 1-25.

Fehr, E. and Gachter, S. "Altruistic Punishment in Humans." *Nature*, 2002, *415*(6868), pp. 137-40.

Fehr, E. and Rockenbach, B. "Detrimental Effects of Sanctions on Human Altruism." *Nature*, 2003, *422*(6928), pp. 137-40.

Fischbacher, U.; Gachter, S. and Fehr, E. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 2001, *71*(3), pp. 397-404.

Gintis, H. "Strong Reciprocity and Human Sociality." *Journal of Theoretical Biology*, 2000, *206*(2), pp. 169-79.

Gintis, H.; Bowles, S.; Boyd, R. and Fehr, E. "Explaining Altruistic Behavior in Humans." *Evolution and Human Behavior*, 2003, *24*(3), pp. 153-72.

Hammerstein, Peter. *Genetic and Cultural Evolution of Cooperation*. Cambridge, Mass.: MIT Press in cooperation with Dahlem University Press, 2003.

Henrich, J. "Cultural Group Selection, Coevolutionary Processes and Large-Scale Cooperation." *Journal of Economic Behavior & Organization*, 2004, *53*(1), pp. 3-35.

Henrich, J. and Boyd, R. "Why People Punish Defectors - Weak Conformist Transmission Can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas." *Journal of Theoretical Biology*, 2001, *208*(1), pp. 79-89.

Henrich, J.; Boyd, R.; Bowles, S.; Camerer, C.; Fehr, E.; Gintis, H. and McElreath, R. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review*, 2001, *91*(2), pp. 73-78.

Johnson, Allen W. and Earle, Timothy K. "The Evolution of Human Societies : From Foraging Group to Agrarian State." 1987.

Kurzban, R. and Houser, D. "Experiments Investigating Cooperative Types in Humans: A Complement to Evolutionary Theory and Simulations." *Proceedings of the National Academy of Sciences of the United States of America*, 2005, *102*(5), pp. 1803-07.

Mahdi, N. Q. "Pukhtunwali - Ostracism and Honor among the Pathan Hill Tribes." *Ethology and Sociobiology*, 1986, *7*(3-4), pp. 295-304.

Masclet, D.; Noussair, C.; Tucker, S. and Villeval, M. C. "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review*, 2003, *93*(1), pp. 366-80.

Ostrom, E.; Burger, J.; Field, C. B.; Norgaard, R. B. and Policansky, D. "Sustainability - Revisiting the Commons: Local Lessons, Global Challenges." *Science*, 1999, *284*(5412), pp. 278-82.

Ostrom, E.; Walker, J. and Gardner, R. "Covenants with and without a Sword - Self-Governance Is Possible." *American Political Science Review*, 1992, *86*(2), pp. 404-17.

Rege, M. and Telle, K. "The Impact of Social Approval and Framing on Cooperation in Public Good Situations." *Journal of Public Economics*, 2004, *88*(7-8), pp. 1625-44.

Stevens, J. R. and Hauser, M. D. "Why Be Nice? Psychological Constraints on the Evolution of Cooperation." *Trends in Cognitive Sciences*, 2004, *8*(2), pp. 60-65.

Wiessner, P. "Norm Enforcement among the Ju/'Hoansi Bushmen - a Case of Strong Reciprocity?" *Human Nature-an Interdisciplinary Biosocial Perspective*, 2005, *16*(2), pp. 115-45.

Yamagishi, T. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology*, 1986, *51*(1), pp. 110-16.

Young, H. Peyton. *Individual Strategy and Social Structure : An Evolutionary Theory of Institutions*. Princeton, N.J.: Princeton University Press, 1998.

2. Community Choice in Social Dilemmas – A "Voting with One's Feet" Approach[5]

2.1. Introduction

Understanding the determinants and the extent of human cooperation is one of the most challenging questions in many fields of science, in particular in economics and political sciences. For example, sustainable use of natural resources, the conservation of the global climate, or the ban of nuclear weapons requires cooperation among members of the "international community", although a defection strategy would be myopically beneficial for individual parties. The conflict between collective and individual interests creates the well-known free-riding problem inherent to social dilemmas (Hardin 1968, Dawes 1980). To enforce and maintain cooperative behavior, humans engage in shaping their social environment, e.g., by creating institutional frameworks. Myriads of laws and regulations have been established throughout the history. Recent examples for such institutions are the "Kyoto Protocol" or the "Organization for Security and Co-operation in Europe (OSCE)". While the former deals with the global greenhouse effect, the latter is dedicated to the promotion of peaceful solutions in international conflicts. However, the success of such institutions or communities largely depends on their acceptance and members' voluntary contributions.

In order to separate between different motives for cooperation and defection, researchers have intensively studied behavior in social dilemmas in controlled laboratory experiments (for overviews see e.g., Ledyard 1995, Croson 1998, Ostrom 1998, Camerer 2003). In sequential social dilemmas remarkable levels of cooperation have been observed, even in interactions with completely anonymous strangers (see e.g., Weimann 1994, Berg et al. 1995, Brandts and Schram 2001, Fehr et al. 1998, 2002, Henrich et al. 2001). In repeated interactions, cooperation is rarely stable and deteriorates to rather low levels towards the end of the interaction period. In a number of recent studies possibilities to punish norm violators have been identified as valuable means to sustain cooperation. Already relatively weak forms of negative sanctions, like symbolic punishment (Masclet et al. 2003) or other forms of social approval (Gächter and Fehr 1999, Rege and Telle 2004) have been shown to be quite effective at least in the short run. If the punishment allows reducing the income of the punished player, it is heavily used, even if punishing incurs costs for the punisher (Yamagishi 1986, Ostrom et al. 1992, Fehr and Gächter 2000, 2002). In addition it has been shown that rewards have a much weaker effect on cooperation (Sefton et al., forthcoming).

A common feature of the experimental studies on sanctioning in social dilemmas is that the framework for the interaction is *exogenously* imposed by the experimenter. However, in reality, humans often *choose* from various existing "institutions" or "communities" that govern the interaction of its members. For example, people choose to work in a certain firm because of the family friendly (flexible) work times or the cooperative atmosphere. Some wealthy citizens move to neighboring states or to other cities because of the low tax levels. These examples raise the question on the "success" of different mechanisms in environments in which members of the society may "vote with their feet" by choosing the institutional mechanisms.

In this paper we study the *endogenous community choice* in a social dilemma situation in which different communities with different institutional frameworks already exist. The

[5] This chapter is based on the working paper "Community Choice in Social Dilemmas – A "Voting with One's Feet" Approach", joint work with Bernd Irlenbusch and Bettina Rockenbach. All authors contributed equally.

novelty of our experimental approach is that members of the society may freely join the community in which they want to interact (i.e., to play the social dilemma). They are free to move back and forth between communities. In particular, a community is not able to "close the doors" and exclude subjects with "undesired behavior" (such as free-riders). Hence, the idea of our setting shows similarities to Tiebout's famous approach according to which people choose a community that matches their preferences best (Tiebout, 1956).

Our analysis of the situation is both theoretically and experimentally. We extend a social dilemma game – as theoretically analyzed by Fehr and Schmidt (1999) and experimentally investigated by Fehr and Gächter (2000) and others – by adding a community choice stage prior to the contribution stage. At the beginning of each repetition each player may choose to join either a community with costly sanctioning possibilities (punishment *or* reward) or a community without any sanctioning options. In each period a player then interacts with all other players who have chosen the same community in that round. Hence, the size and the composition of each community are endogenously determined. This allows us to study the initial and the over-time acceptance of different communities using different mechanisms and their ability to sustain cooperation in social dilemma situations.

Since very recently, a growing number of studies emphasize the importance of the endogenous choice of the interaction parameters in social dilemma situations. Several studies experimentally investigate the endogenous choice of the "partners" (Ehrhart and Keser, 1999, Hauk and Nagel, 2001, Coricelli et al., 2004, Page et al. 2005) by modeling a choice stage previous to the interaction.[6] Another strand of studies is more similar to our approach. They typically consider situations in which a group of individuals decides by vote upon the institutional rules before interacting repeatedly with each other in a *fixed* composition. While in Sutter et al. (2006) inexperienced subjects decide between different institutions by unanimity vote, in Ostrom et al. (1992) and Ertan et al. (2005) subjects interact in alternative institutions before they decide (by majority vote) whether and to what extent punishment shall be present in a further interaction. Botelho et al. (2007) let subjects vote for a final period of play after experiencing different mechanisms. Gürerk et al. (2006) investigate the choice between a sanctioning institution and a non-sanctioning institution with the endogenous choice mechanism as used in this study. But in that study, subjects can only opt for a sanctioning institution in which *both* mechanisms – reward and punishment – are possible. Some other studies investigate endogenous group formation, through voting with possible "mergers" between groups and the power to exclude individuals from groups Charness and Yang (2007) or with various entry/exit options Ahn et al. (forthcoming).

To obtain a benchmark for our experimental findings we theoretically investigate possible equilibria and the incentives to choose a particular community. If the choice is between a non-sanctioning community (NSC) and a community with punishment possibilities (PuC), we find that equilibria with positive contributions occur under very restrictive conditions and do almost never exist if punishment causes relatively mild harm. The reason lies in the fragility of cooperative communities. They attract free-riders who easily destroy the cooperation. Nevertheless, we show that stable constellations with positive contributions are possible. If the choice is between NSC and a community with reward possibilities (ReC), equilibria with positive contributions are rare since these equilibria require a relatively great fraction of players to who must bear considerable costs to reward all other community members.

[6] Partner selection and its effects on behavior is also explored in other contexts, e.g. market interactions (Kirchsteiger et al., 2005, and Brown et al. 2004), and networks (Riedl and Ule 2003).

Two remarkable experimental results are observed. First, having the choice between NSC and PuC, initially, more than two-thirds of the subjects decide to interact in NSC, but over rounds the proportion of the subjects in this community steadily decreases with an almost complete extinction towards the end of the experiment. Second, the majority of subjects who initially choose PuC, heavily punish free-riders and achieve almost full cooperation. This cooperation is surprisingly stable and it even continues when the interacting group becomes large – in fact, in the end almost all subjects join PuC. Interestingly, a choice between ReC and NSC does not lead to such a clear picture. Over time, in both communities a decay of cooperation is observed and subjects move back and forth between the two communities with a more or less constant fraction of the total population in both communities.

2.2. A simple model of community choice

In this section we present a simple model that allows us to study the community choice with a "voting with one's feet" approach. In the *community choice game*, players choose between two communities before interacting in a social dilemma situation with others who choose the same community. Three different communities are considered: the non-sanctioning community (NSC) resembles the standard voluntary contribution mechanism. In the punishment community (PuC) players may punish other players; in the reward community (ReC) players may reward others. Specifically, in PuC and ReC, players have the possibility to influence other members' payoffs after having observed their individual contributions. We focus on two sets on alternatives: first, a situation in which the choice is between NSC and PuC. This is called the PUN treatment. In the other situation the choice is between NSC and ReC to which we will refer as the REW treatment. We consider a three stage game consisting of a "voting with one's feet" stage (S0), a voluntary contribution stage (S1) and a sanctioning stage (S2). With respect to stages S1 and S2 our game is analogous to the game analyzed by Fehr and Schmidt (1999); we augment their design by the community choice stage S0.

2.2.1. Community choice stage

In the community choice stage S0, each of the M players from the population individually chooses in which community he or she wants to interact with other players. The community choice is not costly for the players. Once all players have completed their community choice, they are informed about the total number of players n_θ with $\theta \in \{N, S\}$ who have chosen each of the two communities, where n_N stands for the number of players choosing NSC and n_S for the number of players choosing PuC or ReC, respectively. The identities of the players, however, are not revealed.

2.2.2. Contribution Stage

In the contribution stage S1, a player i interacts solely with those players that have opted for the same community. Each player i is endowed with y monetary units and may contribute g_i $(0 \le g_i \le y)$ to a joint project. Players decide simultaneously on how much they contribute. The amount not contributed remains at the private disposal of the player.[7] The sum of all contributions is multiplied by the *productivity factor R*, that is the total rate of return for the whole community from each token invested into the joint project. The resulting amount is equally distributed among all community members, independent of their individual contribution. Hence, each player receives a *marginal per capita return* (MPCR) a_θ from his/her own contribution. In order to give smaller communities the possibility to be as

[7] If only a single player joins a community, no joint project can be created and the total endowment of the player is automatically transferred to her private account. Therefore this player has no decision in stages S1 and S2.

productive as larger ones, we keep the productivity constant, i.e., $R = n_\theta a_\theta$ with $1/n_\theta < a_\theta < 1$. Thus, independent of the community size, the return from the joint project is always the same if all members contribute their whole endowment. The consequence of holding *R* constant is that a_θ decreases with an increase in n_θ.[8] However, holding *R* constant guarantees that the payoff per player does not vary with n_θ, if each community member contributes $g_i = y$. After all players have taken their contribution decisions, they are informed about the individual contribution of each member in their own community.

2.2.3. Sanctioning Stage

At the beginning of the sanctioning stage S2 each player receives z additional monetary units independent of her affiliation as well as her contribution in S1. Players in PuC may use the tokens to punish (or to reward in ReC) other members of the own community. Hence, all players are equipped with the same sanctioning possibilities, independent of their contributions in S1. Providing players in NSC with the same additional endowment eliminates incentives to choose the sanctioning community just for the extra "sanctioning tokens". For the members of NSC, the active part of the game ends here. The total monetary payoff of player *i* in NSC is

$$(1) \qquad x_i^{NSC} = \left(y - g_i + a_N \sum_{j=1}^{n} g_j\right) + z\,.$$

In PuC (ReC) all players simultaneously decide whether or not to punish (reward) other members of their community. Player *i* can punish (reward) community member *j* by assigning punishment (reward) tokens t_j^i to *j*. Each token assigned by player *i* to player *j* incurs a cost of $c < 1$[9] token for player *i* and reduces (increases) the payoff of player *j* by one token. In total each player may assign up to z tokens. Let T^i denote the amount of tokens that player *i* assigns and T^{-i} denote the amount tokens that player *i* receives from the other members of her community. The total monetary payoff of player *i* in PuC results in

$$(2) \qquad x_i^{PuC} = \left(y - g_i + a_S \sum_{j=1}^{n_S} g_j\right) + \left(z - cT^i - T^{-i}\right).$$

The total monetary payoff of player *i* in ReC results in

$$(3) \qquad x_i^{\mathrm{ReC}} = \left(y - g_i + a_S \sum_{j=1}^{n_S} g_j\right) + \left(z - cT^i + T^{-i}\right).$$

The expressions in parentheses represent the stage payoffs of S1 and S2 respectively. At the end of the game all players are informed about all other players' contributions, their sanctioning tokens assigned, their sanctioning tokens received and their resulting total payoff.

---

[8] Isaac and Walker (1988b) examine the effects of different MPCRs in public good experiments. They find significantly more free-riding behavior in their low MPCR treatment (0.30) than in the high MPCR condition (0.75). Of course this observation emerges from a static comparison between treatments and not from a dynamic change of the MPCR within a population.

[9] For punishment, $c < 1$ reflects the widely accepted assumption, that in general punishing someone is less costly than being punished and proven effective elsewhere (compare Abbink et al. 2000, Fehr and Gächter, 2002, or Andreoni et al., 2003). With $c < 1$ the punisher can reduce the absolute "inequality" in payoffs to his own disadvantage since the punishment action reduces the income of the punished player more than the own income is diminished. For rewards, $c < 1$ implies that rewarding is efficiency increasing.

2.3. Theoretical Analysis

In this section we first analyze the community choice game assuming that *all* players solely maximize their own monetary payoff and do not care for others' payoffs. In a second step, we investigate the community choice assuming that players may care also for others' payoffs. In particular, we will base our considerations on the inequality aversion model proposed by Fehr and Schmidt (1999) who assume that there are players who dislike inequitable outcomes in the sense that they are ready to forego some monetary payoff to reach more equitable outcomes.

2.3.1. Equilibria with purely money maximizing subjects

While full contribution of all players would be socially optimal because $n_\theta a_\theta > 1$, for money maximizing players in NSC contributing zero is the dominant strategy: Since $a_\theta < 1$, it is in the material self interest of each player to withhold contributions. If it is common knowledge that all players aim at maximizing their individual monetary payoff, zero contributions by all players constitute the only (Nash) equilibrium. This is true independent of the community size. In equilibrium each player earns $y + z$ tokens. However, the joint payoff of the community is maximized if each member contributes her entire endowment. In this case each member's payoff rises to $n_\theta a_\theta y + z$ tokens, independent of the community size.

In PuC (ReC), no purely money maximizing player will take the opportunity to punish (reward) in S2 since it is costly to do so. Applying backward induction, rational players foresee this and do not contribute in S1. Hence, independent of the community size in PuC (ReC), in the subgame-perfect equilibrium of the one-shot game, players do not contribute and do not sanction. In this case the total payoff is $y + z$ tokens, which is the same as rational and purely money maximizing players obtain in NSC. Again, each player's payoff would be $n_\theta a_\theta y + z$ tokens if all community members fully cooperated. In ReC, however, the joint community payoff is maximized if each community member contributes the entire endowment and additionally assigns all reward tokens to fellow community members. In this case members' expected/average period payoff would be $n_\theta a_\theta y + (z / c)$ tokens.

In a world of money maximizing rational actors, in both PUN and REW, a player is indifferent between the two communities (i.e., between NSC and PuC, and between NSC and ReC, respectively) because in each of them in equilibrium the identical payoff of $y + z$ will be achieved.

2.3.2. Equilibria with inequality aversion

In this section we analyze the community choice game under the assumption of inequality aversion as modeled by Fehr and Schmidt (1999). Fehr and Schmidt (henceforth FS) assume that there exist at least some players whose utility not only depends on the monetary payoff, but also on the degree and the nature of inequality in the payoff allocation. Specifically, an "inequity averse" player *i* weights inequality in payoffs to her disadvantage with a parameter α_i and inequality in payoffs to her advantage with a parameter β_i. If $x = (x_1,...,x_{n_\theta})$ denotes the vector of the individual monetary payoffs of all community members, player *i*'s utility is described by

$$(4) \qquad U_i(x) = x_i - \alpha_i \frac{1}{n_\theta - 1} \sum_{\substack{j=1 \\ j \neq i}}^{n_\theta} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n_\theta - 1} \sum_{\substack{j=1 \\ j \neq i}}^{n_\theta} \max\{x_i - x_j, 0\} .$$

FS assume that both forms of inequality cause disutility, but advantageous inequality has a lower weight than disadvantageous inequality, which implies $\alpha_i \geq \beta_i$. It is also reasonable to restrict the parameter for advantageous inequality by $\beta_i < 1$ since $\beta_i \geq 1$ would imply that a player is ready to burn one dollar in order to decrease the inequality to his advantage relative to another player whose payoff is *less*. On the other hand, there is no upper bound on α_i.

Following FS we suppose that there exist two types of players. The distinctive characteristic of both types is whether a player is happy to cooperate if others cooperate or not. In general, by contributing one unit, a player i generates an individual monetary benefit of a_θ (MPCR). If others contribute more than i, player i experiences (non-monetary) disutility from advantageous inequality. This means, by increasing the own contribution by one unit, player i can reduce own disutility from advantageous inequality by β_i (assuming $g_j = g > g_i$ for all $j \neq i$). The sum of the monetary and the non-monetary benefit $a_\theta + \beta_i$ is therefore an upper bound for the total return from contributing one extra unit. If this return exceeds the monetary benefit of keeping the unit, i.e., if $a_\theta + \beta_i \geq 1$, then player i will increase own contribution. Following FS, we classify such a player as a *conditional cooperator* because her increase in contribution (cooperation) is triggered by other players who contribute more. If the above condition is not fulfilled for player i, i.e., $a_\theta + \beta_i < 1$, then player i's dominant strategy is to contribute zero to the joint project (see Proposition 4 (a) of FS). Such a player is called a *free-rider*. Because in our model $R = n_\theta a_\theta$, player i is a conditional cooperator if and only if

$$(5) \qquad (R/n_\theta) + \beta_i \geq 1 \text{ and}$$

a free-rider[10] if and only if

$$(6) \qquad (R/n_\theta) + \beta_i < 1 \ .$$

FS apply the inequality aversion model to a public goods game with and without punishment possibilities in a static context. They show that in the absence of punishment (separating) equilibria exist in which a group of conditional cooperators contribute positive amounts while free-riders remain from contribution if the contributors do not suffer too much from disadvantageous inequality.[11] In the presence of punishment possibilities equilibria with positive contributions (of all players) may also exist if (at least) some players suffer sufficiently from disadvantageous inequality.[12]

In the following, we apply the inequality aversion model to our social dilemma context with endogenous choice. Since we allow for the choice *between* the two communities, the number of members and thus the MPCR may vary *within* a community. We assume that the players only derive disutility from inequality with regard to the members of their own community and are not concerned about inequality with regard to the members of the other community. In a first step we will analyze the equilibria for the different communities under the assumption that players are not allowed to choose between communities. In a second step we extend our analysis by allowing for community choice.

[10] Note that "conditional cooperator" and "free-rider" are *names* for different *player types*. In particular, as we will see later in the analysis, a free-rider does not necessarily always free-ride on contributions and a conditional cooperator does not necessarily always contribute.
[11] cf. FS, Proposition 4 (c), p. 839.
[12] cf. FS, Proposition 5, p. 841.

2.3.3. Equilibria in the non-sanctioning community (NSC)

Let k_N ($0 \le k_N \le n_N$) be the number of players with $(R/n_\theta) + \beta_i < 1$, i.e. the number of free-riders. In their Proposition 4 (p. 839), FS identify conditions for the equilibria in a social dilemma game (public goods) without sanctioning possibilities. In part (b) of the proposition, they show that a unique equilibrium exists in which all players withhold contributions if k_N is sufficiently high. Applied to our model, this means that there exists a unique equilibrium with zero contributions in NSC if $k_N > R/2$ (see Corollary 1 in the appendix)[13]. This means that if there are "too many" free-riders in NSC all members of NSC defect.

In Proposition 4 (c) FS show that a (separating) equilibrium may exist in which conditional cooperators contribute a positive amount while free-riders remain from contributing, if conditional cooperators do not suffer too much from disadvantageous inequality. In our model such equilibria may exist if $k_N \le R/2$ and $k_N(n_N - 1) < \frac{a_N + \beta_i - 1}{\alpha_i + \beta_i}$ is satisfied for all conditional cooperators. If all members of NSC are conditional cooperators equilibria with $g_i \ge 0$ also exist (see Corollary 2 in the appendix).

Summary for the equilibria in NSC
For $R = 1.6$, a parameter frequently used in experiments (e.g. Fehr and Gächter, 2000, 2002) as well as in our experiment the considerations above imply that the existence of already one single free-rider in NSC prevents cooperation, i.e., $k_N > R/2$ is satisfied for all $k_N \in \{1,...,n_N\}$. Thus, in our setting, a cooperation-equilibrium in NSC is only possible if all members are conditional cooperators.

2.3.4. Equilibria in the punishment community (PuC)

FS show in their Proposition 5, (p. 841) that in a community with punishment possibilities, a pooling equilibrium (in contributions) is possible (i.e., all members contribute to the joint project) if there are at least some conditional cooperators who sufficiently care about inequality to their disadvantage and thus are ready to punish others who contribute less. We apply the results of FS to our model and extend them in preparation of the analysis of the equilibria under community choice. Following backward reasoning, we first analyze the punishment behavior and then proceed with the contributions.

Suppose that the community is composed of n_S players, conditional cooperators and free-riders, and assume that all players contribute the same amount. Can this be an equilibrium or is it profitable to deviate to a lower contribution? Without any punishment, a deviator would obtain a higher monetary payoff than all conditional cooperators and thus cause disutility from disadvantageous inequality for the cooperators. By punishing the deviator, conditional cooperators reduce the deviator's monetary payoff and thereby decrease their disutility from inequality. The disutility from inequality becomes zero if the punishment exactly equalizes the monetary payoffs of the conditional cooperators and the deviator (Lemma 1). Whether this punishment level will be actually chosen in equilibrium depends on the costs of punishment. Loosely speaking they have to be low enough not to eat up the entire utility gain from equalizing payoffs (Proposition 1). If there is a positive number of conditional cooperators who are ready to punish, equilibria with positive contributions exist (Proposition 2). In the absence of those players no equilibrium with $g_i > 0$ exists, unless the population consists of

[13] Due to space restrictions some corollaries and all proofs are postponed to appendix.

only conditional cooperators (Proposition 3). Obviously, the existence of an equilibrium with positive contributions requires a "sufficient number" of conditional cooperators, who are ready to punish. Corollary 3 formulates conditions under which such players may be present and shows that these restrictions are rather strict.

Lemma 1 Consider a community of n_S players who contribute $g > 0$ to the joint project. If a deviator i with a lower contribution $g_i < g$ is punished by a fraction $n_S' \leq n_S - 1$ of all other players, the punishment level that equalizes the monetary payoffs of the n_S' punishers and the deviator is $\gamma^* = \frac{g - g_i}{n_S' - c}$.

Proposition 1 Consider a community of n_S players who contribute $g > 0$ to the joint project. A deviator i who deviates to a lower contribution $g_i < g$ will be punished by $n_S' \leq n_S - 1$ players with the punishment level γ^* if and only if

$$(7) \qquad c < \frac{\alpha_j}{(n_S - 1)(1 + \alpha_j) - (n_S' - 1)(\alpha_j + \beta_j)}$$

holds for each of the n_S' punishers ($1 \leq j \leq n_S'$). A conditional cooperator who additionally satisfies (7) is called a *conditional cooperative enforcer* (shortly: *enforcer*). Thus as long as the punishment cost c is sufficiently low it does not pay to deviate from the punishment level γ^*, i.e., all enforcers punish with γ^*.

We have shown that there exists a punishment level γ^* which equalizes the payoffs of a contribution deviator and the punishers (Lemma 1) and that there exist players who are ready to act as punishers ("enforcers") if the cost of punishment c is low enough (Proposition 1).

Proposition 2 Consider a community of n_S players with $n_S' > 0$ enforcers satisfying (7). Then there exists a symmetric equilibrium in PuC in which all players contribute g to the joint project and no punishment occurs on the equilibrium path.

Intuitively, this means that if there are a sufficient number of players who are ready to punish the deviators then a cooperation equilibrium with positive contributions exists in which *all* players contribute the same amount to the joint project on the equilibrium path. Hence, on the equilibrium path, there is no need for punishment.

Proposition 3 Let k_S denote the number of free-riders in PuC. Consider a community of n_S players. Suppose that there are at least $\frac{n_S - 1}{2n_S} R$ free-riders and no enforcers satisfying (7) (i.e., $n_S' = 0$). Then in the unique equilibrium all players contribute $g_i = 0$.

If there are no enforcers who threaten to punish the non-contributors the situation in PuC is "equivalent" to the situation in NSC without sanctioning possibilities: if there are a sufficient number of free-riders k_S in PuC, then cooperation is impossible and $g_i = 0$ is played in the unique equilibrium.

As can be seen from Proposition 1, the prevalence of cooperation in PuC depends on condition (7), in particular on the community size n_S, the number of enforcers n_S', and their inequality parameters α_j and β_j. A closer look reveals that the conditions are relatively strict and interesting implications can be deduced.

Corollary 3 In PuC, an enforcer can only exist if $(n_S - n_S') < 1/c$, i.e., if the number of "non-enforcers" is strictly lower than the reciprocal value to the cost of punishing somebody by one unit.

From Corollary 3 it follows that if $c = 1/3$ (as in Fehr and Gächter, 2002 as well as in our experiment), then in the presence of one free-rider the total population can "afford" at most one non-enforcer in an equilibrium with positive contributions, no matter how large the community size is and how many enforcers there are. The intuition for this astonishing implication may be stated as follows: By investing c in punishment, an enforcer reduces the monetary payoff of the deviator by exactly one unit, hence the inequality between the payoffs of the enforcer and the deviator decreases by $(1-c)$, i.e., the enforcer's disutility from being worse off than the deviator decreases exactly by $\frac{\alpha_j}{n_S - 1}(1-c)$. At the same time, the enforcer creates a payoff inequality of c unit with respect to each non-enforcer who does not deviate. This means that the enforcer suffers from a disutility $\frac{\alpha_j}{n_S - 1}c$ with respect to each non-enforcer; in sum $\frac{\alpha_j}{n_S - 1}(n_S - n_S' - 1)c$. For punishment to be profitable for the enforcer, the utility gain from punishing must outweigh the disutility with respect to the non-enforcers, i.e., $\frac{\alpha_j}{n_S - 1}(1-c) > \frac{\alpha_j}{n_S - 1}(n_S - n_S' - 1)c$. This condition is equivalent to what Corollary 3 proposes: $(n_S - n_S') < 1/c$.

From Corollary 3 it follows: if $c = 1/3$ then a cooperation equilibrium cannot exist if there are three (or more) free-riders. In other words, even thousands of players who highly dislike disadvantageous inequality could not "handle" these three free-riders. Note that in an equilibrium with positive contributions a single enforcer may be sufficient if this enforcer has a sufficiently high α_j and $(n_S - n_S') < 1/c$.

Summary for the equilibria in PuC
We find that in the presence of at least one free-rider, equilibria with positive contributions can only exist if the number of non-enforcers is strictly smaller than $1/c$ and there is at least one enforcer who cares sufficiently about disadvantageous inequality. For $c = 1/3$, as in our setting, the presence of two free-riders implies that all other players must be enforcers for a cooperation equilibrium to exist. If all members of PuC are conditional cooperators, equilibria with positive contributions may also exist.

2.3.5. Equilibria in the reward community (ReC)

In this section we extend the analysis of FS by investigating the equilibria under inequality aversion in a social dilemma game with reward possibility. While the analysis of PuC was an adaptation of the considerations of FS to our case with varying community size and MPCR, the following analysis of ReC has no master copy in FS.

First we will show that a (separating) equilibrium with different contributions for free-riders and conditional cooperators does not exist (Lemma 2 and Proposition 4). Then, we will demonstrate that a pooling equilibrium in which *all* players contribute the same amount to the joint project exists, if there are a sufficient number of conditional cooperators who are ready to reward the contributors (Proposition 5).

Lemma 2 Consider a community of n_S players. Suppose that n_S' players contribute $g > 0$ to the joint project whereas $n_S - n_S'$ players refrain from contributing. If the n_S' contributors reward each other, then the reward level that equalizes the payoffs of all n_S players is $\rho^S = \frac{g}{(1-c)(n_S'-1)}$.

Players who contribute to the joint project obtain not only less monetary payoff but also suffer from disadvantageous inequality with respect to the free-riders. However, since rewarding increases the payoff of the rewarded player, contributors may reward each other and increase their monetary payoffs until their payoffs are equal to the free-riders' payoffs. Doing so, contributors can also eliminate the disutility to their disadvantage from the payoff inequality.

When searching an equilibrium in which conditional cooperators and free-riders differ in their behavior, it is obvious to look at cases in which conditional cooperators contribute and free-riders do not contribute. In Corollary 1, it was already discussed that without any sanctioning options this behavior is not part of an equilibrium play for the parameters frequently used in social dilemma games. The question is whether the presence of the reward mechanism may stabilize different contribution levels of conditional cooperators and free-riders. Proposition 4 shows that this is not the case.

Proposition 4 There is no separating equilibrium in which $n_S' \leq n_S$ conditional cooperators contribute $g > 0$ to the joint project and reward each player who contributes g with $\rho^S = \frac{g}{(1-c)(n_S'-1)}$ while the remaining $n_S - n_S'$ players do not contribute and do not reward any player.

The intuition for Proposition 4 is as follows: We have shown that players who contribute to the joint project may equalize their payoffs to free-riders' payoffs (Lemma 2). This situation is, however, not stable, and thus cannot be part of an equilibrium since each free-rider (who knows the rewarding strategy of the conditional cooperators) has the incentive to contribute and thus *appear* as a conditional cooperator in order to be rewarded by the conditional cooperators. This "deviation" of free riders is profitable and creates the instability of the separating behavior.

Does a pooling (in contributions) equilibrium exist, in which all players contribute to the joint project but only the conditional cooperators do reward each contributor while the free riders do not reward?

Proposition 5 Consider a community of n_S players who *all* contribute $g > 0$ to the joint project. Suppose that a group of $2 \leq n_S' \leq n_S$ players is ready to reward each player with the reward level ρ^* who contributes $g > 0$ while the remaining $n_S - n_S'$ players do not reward any other player. The n_S' players will reward each player who contributes $g > 0$ if and only if

$$(8) \qquad c < \frac{\beta_j(n_S'-1) - \alpha_j(n_S - n_S')}{(n_S-1)[n_S - 1 - \beta_j(n_S'-1) + \alpha_j(n_S - n_S')]}$$

holds for each of the n_S' players. A conditional cooperator who also satisfies (8) is called a *reward provider*.

Proposition 5 states that as long as the rewarding cost c is sufficiently low, *reward providers* will reward all contributors although they know that some of the contributors will not allocate rewards. However, if the cost of rewarding is too high, the reward providers may have an incentive to deviate by lowering their reward level. Does this imply that under the assumption of (8) we may observe an equilibrium in which all players contribute $g > 0$ to the joint project and reward providers reward each contributor?

Proposition 6 Consider a community of n_S players with $n_S' \geq 2$ reward providers. Then there exists a pooling equilibrium (in contributions) in ReC in which all players contribute g to the joint project and "reward providers" reward contributors with ρ^* on the equilibrium path.

The intuition for the proof of Proposition 6 is as follows: In Proposition 4 (see also the proof in Appendix) we have shown that given the reward level is sufficiently high (i.e., equalizing contributors' and free-riders' payoffs) then each free-rider has an incentive also to contribute to the joint project in order to be rewarded. Provided that rewarding is sufficiently "cheap" and there are not "too many" free-riders – proposition 5 shows that a "reward provider" is ready to reward all contributors (with any ρ^*) even if there are some players who only contribute but not reward. Hence if reward providers choose the sufficient reward level, "free-riders" will contribute (but not reward). Thus, a cooperation equilibrium with positive contributions exists in which all players contribute the same amount to the joint project on the equilibrium path and "reward providers" reward each contributor.

Corollary 4 (a) If all players in ReC are conditional cooperators and $n_S \geq 3$, then (8) is satisfied. This means that there exists an equilibrium as described in Proposition 6. **(b)** If ReC consists of conditional cooperators and free-riders, then $n_S \geq 4$ has to be satisfied in order to achieve an equilibrium as described in Proposition 6.

Corollary 4 (a) highlights that if all players are conditional cooperators then for almost all community sizes it is true that an equilibrium with positive contributions and positive rewards exist. Only in case of $n_S = 2$, there exist parameters for which an equilibrium with positive contributions exist, however not necessarily an equilibrium with positive contributions *and* positive rewards. Part (b) of the Corollary 4 states that in the presence of free-riders for $n_S \leq 3$ an equilibrium as described in Proposition 6 never exists.

Corollary 5 Consider a community of n_S players. Suppose that there are no reward providers satisfying (8), i.e., $n_S' = 0$, and at least $\frac{n_S - 1}{2n_S} R$ free-riders. Then in the unique equilibrium all players contribute $g_i = 0$.

If there are no reward providers then the situation in ReC is analogous to the situation in NSC, i.e., as if there were no possibilities to reward. This means, in particular, if there is a sufficient number of free-riders in ReC cooperation is impossible and $g_i = 0$ is the unique equilibrium.

Summary for the equilibria in ReC
We find that in ReC, separating equilibria with different contribution levels for different types do not exist (Proposition 4). On the other hand, in ReC, pooling equilibria (in contributions) with positive contributions are possible if there are a sufficient number of conditional cooperators who are ready to reward all contributors (Proposition 5). Cooperation equilibria may also exist if all players are conditional cooperators.

We completed the analysis of the equilibria that can be played *within* different communities. The next section investigates the implications of the considerations above in our setting.

2.3.6. Implications for our setting

The predictions of our model depend on the distribution of the inequality parameters in the population. In the following, we investigate how likely the equilibria are, given specific distributions. In particular, we are interested to find out the different likelihoods of the cooperation equilibria in NSC, PuC, and ReC. To compute these likelihoods, we use two distributions of the inequality aversion parameters: the first distribution is the one used in FS (Table III on p. 844; henceforth, we refer to this distribution as the "FS distribution"). The FS distribution *assumes* an upper limit for the inequality parameter $\beta_i = 0.6$. For our setting, this means that cooperation equilibria cannot exist for $n_\theta \geq 5$ in any community since the condition for being a conditional cooperator $(R/n_\theta) + \beta_i \geq 1$, which is the "basic" condition for any cooperation equilibrium in our model, is never satisfied. Hence, using the FS distribution, for all communities $n_\theta \geq 5$, our model predicts always defection, independent of the size and the composition of the communities.

Fortunately, there exists a growing literature on the "other-regarding" preferences. In one of these studies Blanco et al. (2007) estimate the inequality preferences of subjects. A novel feature of this study is that it reports the joint distribution of the inequality parameters which is necessary to predict the behavior, especially in PuC and ReC. While the estimated α_i's by Blanco et al. (2007) are mainly in line with the distribution of the same parameter by FS, the β_i's found by Blanco et al. (2007) tend to be larger than the ones assumed in FS.[14] In particular, Blanco et al. (2007) report that 11% of the subjects have $\beta_i > 0.8\overline{3}$ and 3% of them have even $\beta_i = 1$. In light of this finding, we adjust the FS preferences slightly, by replacing the maximal value $\beta_i = 0.6$ in FS with $\beta_i = 0.87$.[15] The reason to choose this specific parameter value is that it reflects the fact that the β_i's found by Blanco et al. (2007) are higher than β_i's by FS but also because it guarantees that (at least) 40% of the population satisfies the condition for being a conditional cooperator, independent of the number of players in the community. This adjustment of the FS parameter distribution enables us to make predictions not only for $n_\theta \leq 4$ but for all community sizes. Table 2.1 depicts the preference distributions adopted by FS as well as the estimations of Blanco et al. (2007).

[14] There is a similar study by Bellemare et al. (2007) that also reports higher β_i's than assumed by FS.
[15] We will refer to this distribution as the "adjusted FS parameters".

Table 2.1: Distribution of inequality aversion preferences

Distribution of α_i's				Distribution of β_i's			
Fehr & Schmidt		Blanco et al.		Fehr & Schmidt		Blanco et al.	
$\alpha_i = 0$	30%	$\alpha_i < 0.4$	31%	$\beta_i = 0$	30%	$\beta_i < 0.235$	29%
$\alpha_i = 0.5$	30%	$0.4 \leq \alpha_i < 0.92$	33%	$\beta_i = 0.25$	30%	$0.235 \leq \beta_i < 0.5$	15%
$\alpha_i = 1$	30%	$0.92 \leq \alpha_i < 4.5$	23%	$\beta_i = 0.6$	40%	$0.5 \leq \beta_i$	56%
$\alpha_i = 4$	10%	$4.5 \leq \alpha_i$	13%				

Equilibria in NSC in our setting

We already have shown that in our setting in NSC equilibria with $g > 0$ are only possible if and only if *all* players are conditional cooperators, i.e., if *all* players satisfy $(R/n_\theta) + \beta_i \geq 1$. How likely is this given both distributions introduced above? Table 2.2 depicts the computed probabilities for the existence of a cooperation equilibrium using the FS distribution as well as the adjusted FS parameters.

Table 2.2: Probability computation for cooperation in NSC

n_N	Probability that a player i satisfies $(R / n_\theta) + \beta_i \geq 1$		Probability for a cooperation-equilibrium in NSC	
	FS distribution	Adjusted FS	FS distribution	Adjusted FS
2	.7	.7	.4900	.4900
3	.4	.4	.0640	.0640
4	.4	.4	.0256	.0256
5	.0	.4	.0000	.0102
6	.0	.4	.0000	.0041
7	.0	.4	.0000	.0016
8	.0	.4	.0000	.0007
9	.0	.4	.0000	.0003
10	.0	.4	.0000	.0001
11	.0	.4	.0000	.0000
12	.0	.4	.0000	.0000

Independent of the distribution used, for $n_N = 2$, the probability for a cooperation equilibrium to exist amounts to 49%. However, for $n_N > 2$ the chances for cooperation sharply decrease. For just three players it is about 6.4% and for four players about 2.6%. As stated above, with FS parameters, cooperation in NSC is never possible for $n_N \geq 5$. With the adjusted FS parameters, however, for communities with five or more members, there exist small but positive probabilities for cooperation. However, for all communities $n_N \geq 5$, the chances for cooperation amount to not more than 1 %. To sum up, the chances for cooperation in the absence of any sanctioning mechanism are extremely low. A moderately high probability for cooperation exists only in case of $n_N = 2$.

Equilibria in PuC in our setting

Our considerations above have shown that a cooperation equilibrium in PuC in our setting exists *either* if all players are conditional cooperators *or* if – in the presence of at least one free-rider – the number of non-enforcers is strictly smaller than $1/c$ and there is at least one enforcer who cares sufficiently for disadvantageous inequality. Since $c = 1/3$ in our setting, the presence of two free-riders implies that all other players must be enforcers for a cooperative equilibrium to exist. The presence of one free-rider implies that at least all but one of the remaining players must be enforcers.

To be able to compute probabilities for cooperation in PuC we follow FS and assume a perfect correlation between α_i's and β_i's. As FS already indicate this assumption is not fully realistic but it simplifies the analysis. The perfect correlation implies for the FS distribution, that all players with $\alpha_i = 1$ or $\alpha_i = 4$ are assumed to have $\beta_i = 0.6$ whereas players with $\alpha_i = 0$ and $\alpha_i = 0.5$ are assumed to have $\beta_i = 0$ and $\beta_i = 0.25$, respectively. To have also a perfectly correlated adjusted FS distribution, we simply replace $\beta_i = 0.6$ with $\beta_i = 0.87$.

For a given community size, there are three categories of compositions of players in PuC for which a cooperation equilibrium may occur: either (a) all players are conditional cooperators (with sufficiently high β_i's) or (b) there is $n_S - n_S' = 1$ one free-rider in PuC; in this case the remaining n_S' players must be enforcers whose lowest must be $\alpha_i \geq 0.5$ for $n_S = 2$ and $\alpha_i \geq 1$ for $n_S \geq 3$ or (c) there are $n_S - n_S' = 2$ free-riders in PuC (or one free-rider and one non-enforcer); in this case the remaining n_S' players must be enforcers with $\alpha_i \geq 4$ for $n_S \geq 3$. Table 2.3 depicts the necessary number of enforcers and their associated inequality parameters for cooperation. In the last two columns of Table 2.3 the computed "total" probabilities for cooperation are depicted. These total numbers include all the probabilities resulting from the three possible categories of cooperation equilibria.

Table 2.3: Probability computation for cooperation in PuC

	Lowest number of enforcers and lower bound of their α_i's for a cooperation-equilibrium in PuC		Probability that a player *i* has sufficiently high α_i and β_i					
n_S	If there is one free-rider: $n_S - n_S' = 1$	If there are one free-rider and one non-enforcer or two free-riders: $n_S - n_S' = 2$	$n_S - n_S' = 1$		$n_S - n_S' = 2$		Probability for a cooperation-equilibrium in PuC	
			FS	Adjusted FS	FS	Adjusted FS	FS	Adjusted FS
2	$n_S' = 1, \alpha_i \geq 0.5$	-	.7	.7	-	-	.9100	.9100
3	$n_S' = 2, \alpha_i \geq 1$	$n_S' = 1, \alpha_i = 4$	.3	.3	.1	.1	.4600	.4600
4	$n_S' = 3, \alpha_i \geq 1$	$n_S' = 2, \alpha_i = 4$	.3	.3	.1	.1	.2008	.2008
5	$n_S' = 4, \alpha_i \geq 1$	$n_S' = 3, \alpha_i = 4$	.0	.3	.0	.1	.0000	.0906
6	$n_S' = 5, \alpha_i \geq 1$	$n_S' = 4, \alpha_i = 4$	.0	.3	.0	.1	.0000	.0402
7	$n_S' = 6, \alpha_i \geq 1$	$n_S' = 5, \alpha_i = 4$	.0	.3	.0	.1	.0000	.0187
8	$n_S' = 7, \alpha_i \geq 1$	$n_S' = 6, \alpha_i = 4$	.0	.3	.0	.1	.0000	.0085
9	$n_S' = 8, \alpha_i \geq 1$	$n_S' = 7, \alpha_i = 4$	.0	.3	.0	.1	.0000	.0037
10	$n_S' = 9, \alpha_i \geq 1$	$n_S' = 8, \alpha_i = 4$	.0	.3	.0	.1	.0000	.0017
11	$n_S' = 10, \alpha_i \geq 1$	$n_S' = 9, \alpha_i = 4$	.0	.3	.0	.1	.0000	.0006
12	$n_S' = 11, \alpha_i \geq 1$	$n_S' = 10, \alpha_i = 4$	.0	.3	.0	.1	.0000	.0003

Independent of the parameter distribution, the probability for cooperation amounts to 20% or more for small communities ($n_S \leq 4$). With FS distribution, cooperation in PuC is not possible for $n_S > 4$ (as in case of NSC). With the adjusted FS preferences, for $n_S \geq 5$, the chances for cooperation amount to 10% or below, for larger communities ($n_S \geq 8$) the probabilities for cooperation even decrease to 0.1% and lower. Hence, we can summarize as the prospects for cooperation in PuC as follows: for small communities there exist a considerable probability for cooperation whereas for large communities the chances for cooperation are extremely low.

Equilibria in ReC in our setting

We found that in ReC, pooling equilibria (in contributions) with positive contributions are possible if there are a sufficient number of reward providers. Cooperation equilibria also exist if all players are conditional cooperators.

In Table 2.4, the lowest number of reward providers (conditional cooperators) that is necessary for a cooperation equilibrium is depicted. As in case of PuC we assume a perfect correlation between α_i's and β_i's. The total probabilities for cooperation are depicted in the last two columns. Independent of the distribution of parameters, for $n_S = 2$, the probability for a cooperation equilibrium amounts to 49%. For small communities, the chances for cooperation are below 10%. With the FS distribution, cooperation in ReC is not possible for $n_S > 4$ (as in case of NSC and PuC). With the adjusted FS parameters, the probabilities remain well below 10% for $n_S \geq 5$. Hence, in our setting, the possibility to reward does not considerably improve the chances for cooperation though – compared to NSC and PuC – for larger communities the prospects for cooperation are slightly better.

Table 2.4: Probability computation for cooperation in ReC

	Lowest number of reward providers and their α_i's for a cooperation equilibrium in ReC		Probability that a player has sufficiently high α_i and β_i				Probability for a cooperation-equilibrium in ReC	
	$\alpha_i = 1$	$\alpha_i = 4$	$\alpha_i = 1$	$\alpha_i = 4$	$\alpha_i = 1$	$\alpha_i = 4$		
n_S			FS	Adjusted FS	FS	Adjusted FS	FS	Adjusted FS
2	$n_S' = 2$	$n_S' = 2$	.6	.6	.1	.1	.4900	.4900
3	$n_S' = 3$	$n_S' = 3$	.3	.3	.1	.1	.0640	.0640
4	$n_S' = 3$	$n_S' = 4$	.3	.3	.1	.1	.0904	.0904
5	$n_S' = 4$	$n_S' = 5$	.0	.3	.0	.1	.0000	.0345
6	$n_S' = 4$	$n_S' = 6$	.0	.3	.0	.1	.0000	.0712
7	$n_S' = 5$	$n_S' = 7$	.0	.3	.0	.1	.0000	.0292
8	$n_S' = 5$	$n_S' = 7$	.0	.3	.0	.1	.0000	.0580
9	$n_S' = 6$	$n_S' = 8$	.0	.3	.0	.1	.0000	.0253
10	$n_S' = 6$	$n_S' = 9$	.0	.3	.0	.1	.0000	.0474
11	$n_S' = 7$	$n_S' = 10$	.0	.3	.0	.1	.0000	.0216
12	$n_S' = 8$	$n_S' = 11$	.0	.3	.0	.1	.0000	.0095

2.3.7. Community choice equilibria

Now we relax the assumption of isolated communities and look for equilibria in an environment in which each member can freely choose between communities. Are there equilibria in which both communities are simultaneously populated, and if so with different contribution levels? Or, are there only equilibria in which one community is depopulated? To approach these questions we assume that in each of both communities an equilibrium is played. We seek for stable allocations where no individual player has an incentive to deviate from his strategy by choosing the other community. As FS, we assume that players have individual parameters for inequality aversion which neither change over time nor are they influenced by the parameters of the other players. However, remember that the MPCR of a community depends on the number of players in that community ($a_\theta = R / n_\theta$). By choosing a community, a player increases n_θ and thus reduces a_θ of that community. As a consequence, a player's "type" may depend on the community choice of the player as well as of the other players. For example, player i may be a free-rider in case i joins NSC because $R / n_N + \beta_i < 1$, whereas the same player may be a conditional cooperator in case i joins PuC

because $R/n_S + \beta_i \geq 1$, or vice versa. Moreover, even other players' types may depend on the choice of player i since their types also depend on a_θ. Hence, when making the community choice, a player has to consider the consequences for equilibrium play. This is a specialty of our design which may appear counterintuitive when not only the parameters α_i and β_i but also the types are understood as hard-wired characteristics of a player. A deeper reflection, however, shows that it is appropriate to consider the types as being variable, because the term $R/n_\theta + \beta_i$ reflects the sum of the monetary payoff and the utility gain from avoiding advantageous inequality. In one community, the monetary payoff R/n_θ may be sufficient to compensate an investment of one token whereas in the other community the MPCR is too low to compensate this investment. Hence, in one community the player appears as a conditional cooperator whereas in the other community the same player acts as free-rider.

To simplify the analysis of the equilibria under community choice, we assume that there is only one level of cooperation in equilibrium in each of both communities which is in particular independent of n_θ. We denote the equilibrium level of cooperation in NSC by $g_N^* > 0$ and the equilibrium level of cooperation in PuC (and ReC, respectively) by $g_S^* > 0$. Yet, we do not make any assumption about the relationship between g_N^* and g_S^*, because Corollary 2, Proposition 2, and Proposition 6 showed that in equilibrium cooperation is possible on various levels $0 \leq g^* \leq y$.

Definition

1. A *defection equilibrium* is an equilibrium in NSC, PuC, or ReC with the properties described in Corollary 1, Proposition 3, and Corollary 5, respectively. In a defection equilibrium all players contribute $g_\theta^* = 0$ on the equilibrium path.

2. A *cooperation equilibrium* is an equilibrium in NSC, PuC, or ReC with the properties described in Corollary 2, Proposition 1, and Proposition 6. In a cooperation equilibrium all players contribute $g_\theta^* \geq 0$ on the equilibrium path.

3. A *community choice equilibrium* is a stable allocation of the players to two communities with the property that a cooperation equilibrium and/or a defection equilibrium is played in each of both communities and no player has an incentive to deviate from his or her community choice.

In the following, we analyze first the community choice equilibria in PUN, i.e., when the choice is between NSC and PuC.

Community choice equilibria in PUN

The following allocations of the players to NSC and PuC and respective contribution levels (as listed in Table 2.5) can be sustained by a community choice equilibrium provided that the indicated conditions are satisfied.

1. Defection equilibrium played in NSC: The number of free-riders in NSC is $k_N \geq \frac{n_N - 1}{2n_N} R$ (see Corollary 1).
2. Cooperation equilibrium played in NSC: In NSC, all players are conditional cooperators (see Corollary 2).

3. Defection equilibrium played in PuC: In PuC, there exists no enforcer ($n_S' = 0$) and there are at least $\frac{n_S - 1}{2n_S} R$ free-riders (see Proposition 3).

4. Cooperation equilibrium played in PuC: In PuC, there exist $n_S' > 0$ enforcers (see Proposition 1).

5. For all players in PuC it must hold: If this player deviates from PuC to NSC then a cooperation equilibrium in NSC is not possible. This means if a player in PuC chooses NSC instead then the number of free-riders after deviation in NSC, k_N^D, satisfies $k_N^D \geq \frac{n_N}{2(n_N + 1)} R$ (see Corollary 1).

6. For all players in NSC it must hold: If this player deviates from NSC to PuC then a cooperation equilibrium in PuC is not possible. If a player in NSC chooses PuC instead then there exists no enforcer ($n_S' = 0$) who satisfies $a + \beta_j \geq 1$ and $c < \frac{\alpha_j}{[(n_S + 1) - 1](1 + \alpha_j) - (n_S' - 1)(\alpha_j + \beta_j)}$ (Proposition 1).

Conditions 1-4 guarantee that behavior in each community is stable. Additionally, we have to make sure that no player has an incentive to switch from one community to the other which is given by conditions 5 and 6. Feasible combinations of these conditions characterize the community choice equilibria. The intuitions for these are provided in the following.

Community choice equilibria a)-d): Only one of the two communities is populated and all players either play a defection *or* a cooperation equilibrium. In either case, no player has an incentive to change to the depopulated community because a single player is not able to contribute to a joint project.

Community choice equilibrium e): Both communities are populated and in both communities a cooperation equilibrium is played with *equal* contributions. By assumption, in the cooperation-equilibrium players' contributions do not depend on the number of players in their community. This means that if a member of one community would deviate from his or her community choice to join the other community, this would not change the contributions (in the "new" cooperation equilibrium) in the other community. Thus, no player in any community has an incentive to deviate from his or her community choice.

Community choice equilibria f) and g): A cooperation equilibrium is played in NSC *and* in PuC, however with different contribution levels. In that case players would have an incentive to deviate from the community with the lower contribution level to the community with the higher contributions, if the cooperation equilibrium would still be played after deviating. Parts f) and g) describe the conditions under which different equilibrium contribution levels constitute a community choice equilibrium, by specifying the conditions that the cooperation equilibrium with the higher contributions is no longer possible *if* a player deviates.

Table 2.5: Community choice equilibria in PUN

	One of the communities is depopulated				Both communities are populated								
	Defect. in NSC	Coop. in NSC	Defect. in PuC	Coop. in PuC	Coop. in NSC & coop. in PuC			Defect. in NSC & Coop. in PuC		Coop. in NSC & Defect. in PuC		Defect. in NSC & defect. in PuC
	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**
Number of players in NSC n_N	N	N	0	0	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1
Contribution level in NSC g_N	0	≥ 0	–	–	≥ 0	≥ 0	> 0	0	0	> 0	0	0
Number of players in PuC n_S	0	0	N	N	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1
Contribution level in PuC g_S	–	–	0	≥ 0	≥ 0	> 0	≥ 0	> 0	0	0	0	0
					$g_N = g_S$	$g_N < g_S$	$g_N > g_S$	$g_N < g_S$	$g_N = g_S$	$g_N > g_S$	$g_N = g_S$	$g_N = g_S$
Conditions												
1. Defection-equilibrium played in NSC	×							×	×			×
2. Cooperation-equilibrium played in NSC		×			×	×	×			×	×	
3. Defection-equilibrium played in PuC			×							×	×	×
4. Cooperation-equilibrium played in PuC				×	×	×	×	×	×			
5. For all players in PuC it must hold: If this player deviates from PuC to NSC then a cooperation-equilibrium in NSC is not possible.							×	×	×	×		×
6. For all players in NSC it must hold: If this player deviates from NSC to PuC then a cooperation-equilibrium in PuC is not possible.						×		×		×	×	×

Community choice equilibria h) and j): Part h) describes the situation that a defection equilibrium is played in NSC *and* a cooperation-equilibrium with $g_S^* \geq 0$ is played in PuC. In that case players would have an incentive to deviate from NSC to PuC, if the cooperation equilibrium in PuC is still possible despite deviation of members of NSC. Hence, to be h) a community choice equilibrium, for all players in NSC it must hold that if this player deviates from NSC to PuC then a cooperation equilibrium in PuC is not possible (anymore). On the other hand, a deviation from PuC to NSC could possibly lead to a (new) cooperation-equilibrium in NSC with higher contributions than in PuC. Thus, it is also required that if a player deviates from PuC to NSC then a cooperation equilibrium in NSC is (still) not possible. A similar line of argument can be applied to part j) of the proposition.

Community choice equilibria i) and k): Part i) describes the situation that a defection equilibrium is played in NSC *and* a cooperation equilibrium with $g_S^* = 0$ is played in PuC. In this case no player has an incentive to deviate from NSC to PuC since $g_S^* = 0$. On the other hand, a deviation from PuC to NSC could possibly lead to a cooperation equilibrium in NSC with $g_N^* > 0$. Hence, to be i) a community choice equilibrium, for all players in PuC holds that a cooperation equilibrium in NSC is (still) not possible if the player deviates from PuC to NSC. A similar line of argument applies to part k) of the proposition.

Community choice equilibrium l): A defection-equilibrium is played in NSC *and* in PuC. This means a deviation from PuC to NSC (and vice versa) could possibly lead to a cooperation-equilibrium in NSC (or in PuC, respectively). In order for this situation to be a community-choice-equilibrium there must be no incentive to deviate in either direction. Thus, for all

players in NSC it must hold that if this player deviates from NSC to PuC then a cooperation-equilibrium in PuC is (still) not possible and vice versa.

Community choice equilibria in REW
The following allocations of the players to NSC and ReC and respective contribution levels (as listed in Table 2.6) can be sustained by a community choice equilibrium provided that the marked conditions are satisfied. Explicitly, the conditions are as follows:

1. Defection-equilibrium played in NSC: In NSC, the number of free-riders is $k_N \geq \frac{n_N - 1}{2n_N} R$ (see Corollary 1).
2. Cooperation-equilibrium played in NSC: In NSC, all players are conditional cooperators (see Corollary 2).
3. Defection-equilibrium played in ReC: In ReC, there exists no reward provider ($n_S' = 0$) and there are at least $\frac{n_S - 1}{2n_S} R$ free-riders (see Corollary 5).
4. Cooperation-equilibrium played in ReC: In ReC, there exist $n_S' > 0$ reward providers (see Proposition 5) and $n_S - n_S'$ free-riders with sufficiently low β_i's (see Proposition 6).
5. For all players in ReC it must hold: If this player deviates from ReC to NSC then a cooperation-equilibrium in NSC is not possible. If a player in ReC chooses NSC instead then the number of free-riders in NSC after deviation, k_N^D, must satisfy $k_N^D \geq \frac{n_N}{2(n_N + 1)} R$ (see Corollary 1).
6. For all players in NSC it must hold: If this player deviates from NSC to ReC then a cooperation-equilibrium in ReC is not possible. If a player in NSC chooses ReC instead then there exists no reward provider ($n_S' = 0$) who satisfies $a + \beta_j \geq 1$ and $c < \frac{\beta_j (n_S' - 1) - \alpha_j (n_S + 1 - n_S')}{n_S [n_S - \beta_j (n_S' - 1) + \alpha_j (n_S + 1 - n_S')]}$ (see Proposition 6).

Analogous to PUN, conditions 1-4 guarantee that behavior in each community of REW is stable. Conditions 5 and 6 make sure that no player has an incentive to switch from ReC to NSC and vice versa. Feasible combinations of these conditions characterize the community choice equilibria. The intuitions for these equilibria are analogous to those provided for PUN.

Table 2.6: Community choice equilibria in REW

	One of the communities is depopulated				Both communities are populated								
	Defect. in NSC	Coop. in NSC	Defect. in ReC	Coop. in ReC	Coop. in NSC & coop. in ReC			Defect. in NSC Coop. in ReC		Coop. in NSC Defect. in ReC		Defect. in NSC & defect. in ReC
	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**
Number of players in NSC	N	N	0	0	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1
Contribution level in NSC	0	≥ 0	–	–	≥ 0	≥ 0	> 0	0	0	> 0	0	0
Number of players in ReC	0	0	N	N	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1	≥ 1
Contribution level in ReC	–	–	0	≥ 0	≥ 0	> 0	≥ 0	> 0	0	0	0	0
Reward level ρ*		≥ 0		≥ 0	≥ 0	≥ 0	≥ 0	≥ 0	≥ 0			
Conditions					$g_N = g_S$	$g_N < g_S$	$g_N > g_S$	$g_N < g_S$	$g_N = g_S$	$g_N > g_S$	$g_N = g_S$	$g_N = g_S$
1. Defection-equilibrium played in NSC	×							×	×			×
2. Cooperation-equilibrium played in NSC		×			×	×	×			×	×	
3. Defection-equilibrium played in ReC			×							×	×	×
4. Cooperation-equilibrium played in ReC				×	×	×	×	×	×			
5. For all players in ReC it must hold: If this player deviates from ReC to NSC then a cooperation-equilibrium in NSC is not possible.							×	×	×	×		×
6. For all players in NSC it must hold: If this player deviates from NSC to ReC then a cooperation-equilibrium in ReC is not possible.						×		×		×	×	×

Community choice equilibria in our setting

In our setting, all community choice equilibria for PUN and REW described above are theoretically possible. However, the probabilities for the existence of the equilibria depend on the (actual) distribution of players' inequality parameters. To calculate the probabilities for these equilibria we conduct two simulations for PUN and REW based on 1 million iterations. In each iteration, we randomly draw α_i's and β_i's for all players according to the FS distribution. For the drawn parameters we verify the equilibrium conditions specified in Table 2.5 and Table 2.6, respectively. Table 2.7 depicts the percentage of cases where conditions for the respective equilibria are satisfied.[16]

Community choice equilibria in PUN can only be found with substantial probabilities in cases in which defection is played in both communities, i.e., in equilibrium i) for low occupations in PuC and in equilibria l), and k) for $n_S = 10$, except for the degenerated cases in which only one community is populated, i.e., equilibria a) and c). A very similar picture emerges in REW. Thus, most likely we have to expect defection in both communities in both treatments of our experiment.

[16] Due to space restrictions we report only the results of simulations basing on the FS parameter distribution. We conducted also simulations with adjusted FS preferences. However, their results are not significantly different.

Table 2.7: Probability calculation for the community choice equilibria in PUN and REW

		One of the communities is depopulated				Both communities are populated							
PUN		Defect. in NSC	Coop. in NSC	Defect. in PuC	Coop. in PuC	Coop. in NSC Coop. in PuC			Defect. in NSC Coop. in PuC		Coop. in NSC Defect. in PuC		Defect. in NSC Defect. in PuC
		a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**
Conditions to be satisfied:		1	2	3	4	2,4	2,4,6	2,4,5	1,4,5,6	1,4,5	2,3,5,6	2,3,6	1,3,5,6
n_S	n_N												
0	12	1.0000	.0000	-	-	-	-	-	-	-	-	-	-
1	11	-	-	-	-	-	-	-	-	-	-	-	-
2	10	-	-	-	-	.0001	.0000	.0000	.0979	.9093	.0000	.0000	.0314
3	9	-	-	-	-	.0001	.0000	.0000	.0456	.4597	.0001	.0002	.5401
4	8	-	-	-	-	.0001	.0000	.0000	.0122	.2012	.0001	.0006	.7981
5	7	-	-	-	-	.0001	.0000	.0000	.0042	.0905	.0001	.0015	.9078
6	6	-	-	-	-	.0002	.0000	.0000	.0023	.0413	.0002	.0040	.9545
7	5	-	-	-	-	.0002	.0000	.0000	.0015	.0187	.0003	.0101	.9710
8	4	-	-	-	-	.0002	.0000	.0000	.0012	.0083	.0004	.0253	.9662
9	3	-	-	-	-	.0003	.0002	.0000	.0034	.0036	.0006	.0638	.9324
10	2	-	-	-	-	.0001	.0000	.0000	.0000	.0000	.3310	.4896	.5103
11	1	-	-	-	-	-	-	-	-	-	-	-	-
12	0	-	-	1.0000	.0000	-	-	-	-	-	-	-	-
REW		Defect. in NSC	Coop. in NSC	Defect. in ReC	Coop. in ReC	Coop. in NSC Coop. in ReC			Defect. in NSC Coop. in ReC		Coop. in NSC Defect. in ReC		Defect. in NSC Defect. in ReC
		a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**
Conditions to be satisfied:		1	2	3	4	2,4	2,4,6	2,4,5	1,4,5,6	1,4,5	2,3,5,6	2,3,6	1,3,5,6
n_S	n_N												
0	12	1.0000	.0000	-	-	-	-	-	-	-	-	-	-
1	11	-	-	-	-	-	-	-	-	-	-	-	-
2	10	-	-	-	-	.0001	.0000	.0000	.3311	.4900	.0000	.0001	.5099
3	9	-	-	-	-	.0000	.0000	.0000	.0006	.0637	.0001	.0003	.9361
4	8	-	-	-	-	.0000	.0000	.0000	.0004	.0255	.0001	.0006	.9739
5	7	-	-	-	-	.0000	.0000	.0000	.0003	.0102	.0001	.0016	.9882
6	6	-	-	-	-	.0000	.0000	.0000	.0002	.0040	.0002	.0040	.9920
7	5	-	-	-	-	.0000	.0000	.0000	.0001	.0016	.0003	.0101	.9883
8	4	-	-	-	-	.0000	.0000	.0000	.0001	.0006	.0004	.0255	.9739
9	3	-	-	-	-	.0000	.0000	.0000	.0000	.0003	.0006	.0640	.9357
10	2	-	-	-	-	.0001	.0000	.0000	.0000	.0001	.3303	.4889	.5110
11	1	-	-	-	-	-	-	-	-	-	-	-	-
12	0	-	-	1.0000	.0000	-	-	-	-	-	-	-	-

2.4. Experimental Design and Procedure

Following the model described in Section 2 our experimental design consists of two treatments that differ with respect to the communities available in the choice set (see Table 2.8).

Table 2.8: Experimental Treatments

Treatment	Co-existing communities	# Members in a community	# Independent observations
PUN	NSC and PuC	$0 \leq n_\theta \leq 12$	8
REW	NSC and ReC	$0 \leq n_\theta \leq 12$	6

In each treatment $N = 12$ players constitute a population that remains constant during the whole session. In each session one of the two games described in Section 2 is repeated over 30 rounds. By randomly reshuffling the presentation ordering on the computer screens we made sure that the identity of the players could not be traced over rounds. Since players can choose between communities $\theta \in \{N, S\}$ the community size n_θ may vary in each round.[17]

[17] In our experiment we consider a special case of a partner design in which not all members of the group necessarily interact in each round. For an investigation of the differences in behavior of strangers and partners in social dilemma situations see e.g. Croson (1996) or Keser and van Winden (2000).

To compare the success of the two communities, the productivity factor R is kept constant and therefore the MPCR a_θ varies with n_θ. As in many other experimental studies we set $R = 1.6$. Table 2.9 shows the corresponding a_θ's in our experimental design dependent on n_θ. It is worth noting that the threat of punishment and the opportunities to be rewarded is likely to be heavier in larger groups, as in total there are more punishment respective reward tokens available.

Table 2.9: Marginal per Capita Return a(n)

Community size n	2	3	4	5	6	7	8	9	10	11	12
Marginal per Capita Return $a(n)$	0.80	0.53	0.40	0.32	0.27	0.23	0.20	0.18	0.16	0.15	0.13

The experiment was conducted in the computerized laboratory *eLab* of the University Erfurt. The subjects were recruited for voluntary participation on campus. All subjects were undergraduate students. Each subject was allowed to participate only once and none had participated in a similar experiment before. The experiment was programmed and conducted with the software *z-Tree* (Fischbacher, 2007). Communication other than via the experimental software was not allowed.[18]

In the experiment a total of 168 subjects participated in 7 sessions with 24 subjects each. Each session contained two independent observations: In total we have 8 independent observations for PUN and 6 for REW. Before starting the experiment the instructions were read aloud to all participants.[19] The subjects were informed about the experimental procedure as well as the number of rounds. One session lasted for about 2 to 2.5 hours. The subjects earned between 15 to 25 Euros and payments were made anonymously at the end of the experiment.

2.5. Results

In this section we report our experimental findings. First we address the initial community choice and contribution behavior. We continue by investigating the development of the community choice, contributions, and efficiency within the communities and between the treatments. Then, we compare the theoretical predictions to our experimental findings. An in-depth analysis of the community choice behavior and its effects on contributions follow. The conclusion section summarizes our main findings and provides a discussion.

2.5.1. Community choice and the contributions in the first period

The initial choice of community differs between settings (see Table 2.10). More than two thirds (68.8%) of the subjects prefer NSC if the alternative is PuC. In contrast, if ReC is the alternative, more than three fourth of the participants opt for the ReC (76.4%).[20] Thus, the first observation supports the unequal choice frequencies between PuC and NSC and exhibits a clear direction, namely the avoidance of the punishment threat. The second observation reveals a clear affinity towards the reward mechanism.

[18] On the influence of communication and (non-binding) promises on contribution behavior in social dilemmas see Isaac and Walker (1988a), Orbell et al. (1988), and Ostrom et al. (1992), Brosig et al. (2002).

[19] A translation of the instruction sheet is given in Appendix. Original instructions were written in German. They are available upon request from the authors.

[20] Binomial tests show that significantly more subjects choose NSC than PuC (p = 0.000) and more subjects prefer ReC to NSC (p = 0.000), respectively. The reported non-parametric statistical tests are two-tailed if not stated otherwise and use the session averages as independent observations.

Table 2.10: Community choice and average contributions in the first period

Treatment	Non-Sanctioning Community (NSC)			Sanctioning Community (PuC / ReC)		
	Percentage of subjects choosing the					
PUN	68.8			31.2		
REW	23.6			76.4		
	Average contributions in…					
	NSC			PuC / ReC		
PUN	7.4			13.2		
REW	2.7			7.9		
	Percentage of contributions in…					
	$g \leq 5$	$5 < g < 15$	$15 \leq g$	$g \leq 5$	$5 < g < 15$	$15 \leq g$
PUN	43.7	42.2	14.1	18.4	27.9	53.7
REW	93.3	6.7	0.0	47.9	35.5	16.6

In both treatments the initial contributions are higher in the sanctioning communities than in the NSCs. This is true for PUN (U-Test, $p = 0.016$) as well as for REW ($p = 0.063$). The highest first period contributions are made by the subjects who join PuC. They contribute on average 13.2 tokens, (66% of the endowment). Initial contributions in ReC are significantly lower than the contributions in PuC (U-Test, $p = 0.003$). More than half of the subjects (53.7%) who choose PuC in the first period contribute 15 tokens or more. In contrast, only 18.4% of subjects in PuC are free-riders, i.e., they contribute 5 tokens or less. In ReC only 16.6% of subjects contribute high while almost half of the subjects who opt for ReC in the first period are free-riders (47.9%). Initially, the contributions in NSC of PUN are clearly higher than the contributions in NSC of REW (U-Test, $p = 0.002$) since almost all subjects (93.3%) who choose NSC in REW are free-riders whereas in NSC of PUN only 43.7% of subjects are free-riders.

2.5.2. Evolution of community choice and contributions

The development of the community choices is displayed in Figure 2.1: Panel a) displays the percentage of subjects in the two communities of PUN in the course of the rounds. Panel b) displays this data for REW. In PUN, the initial low acceptance of PuC increases steadily with an almost complete extinction of NSC towards the end of the interaction. In the final rounds, more than 90% join PuC.[21] In REW, in contrast, the number of subjects in both communities is rather constant over the whole experiment.[22]

[21] In each of the eight independent observations of PUN the number of members in NSC has a negative trend over time (determined by Spearman-rank-correlation coefficients). Hence a binomial test with these correlation coefficients clearly rejects that a negative trend is as likely as a positive trend ($p = 0.016$).

[22] For one of the six independent observations of REW the number of members in NSC has a negative trend over time, in the remaining five observations no trend can be found.

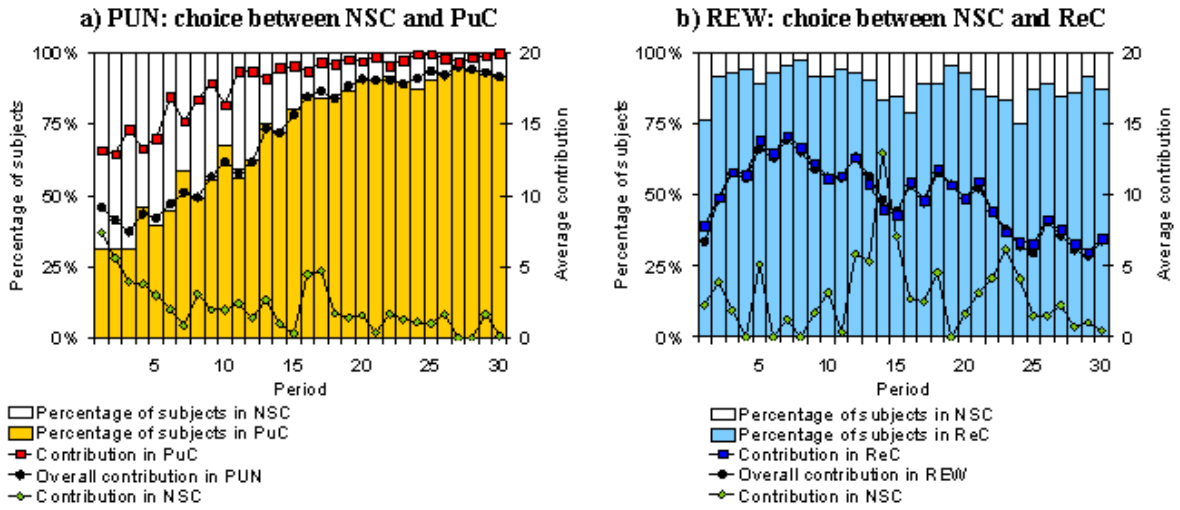


Figure 2.1 Community choice over periods

One of the most important questions regarding social dilemmas concerns the induced contribution behavior. Table 2.11 summarizes the contribution levels in the different treatments and communities, over all 30 periods as well as separately for the first and the second half. Additionally contributions of the first and the last round are shown. Averaged over 30 periods, the contributions in PUN are significantly higher than in REW (U-Test, p = 0.043).[23]

Table 2.11: Average Contributions

Treatment	Periods 1-30		Periods 1-15		Periods 16-30		First Period		Last Period	
PUN	14.5		10.9		18.1		9.2		18.3	
REW	9.8		11.2		8.3		6.7		6.8	
					Communities					
	NSC	PuC/ ReC	NSC	PuC/ ReC	NSC	PuC/ ReC	NSC	PuC/ ReC	NSC	PuC/ ReC
PUN	3.6	18.7	3.6	17.0	3.4	19.5	7.4	13.2	0.2	20.0
REW	5.3	10.1	7.0	11.5	4.4	8.6	2.7	7.9	0.6	7.0

Within each treatment NSC performs worst; it exhibits significantly lower contributions than the alternative community that allows an influence of the other subjects' payoffs. A Wilcoxon matched-pairs test show that the contributions in PuC are significantly higher than in NSC in PUN (p = 0.008) whereas contributions in ReC are not significantly higher than NSC in REW (p = 0.125).[24] A comparison of all communities shows that the highest contributions are observed in PuC, where, on average, subjects contribute 93.5% of their endowment.

As Figure 2.1, Panel a) shows the members of PuC reach almost full cooperation towards the end of the interaction period although the communities are quite large in the final third (about 9 to 12 players) and as a consequence the MPCR is relatively low.[25] Over time, we observe a

[23] This observation is well in line with the results of other studies, e.g., Sefton et al. (forthcoming).
[24] Actually, in REW, in four of five observations, contributions are considerably higher in ReC than in NSC. In only one observation, subjects slightly contribute more in NSC than in ReC. In the sixth observation, all subjects but one always opt for ReC, thus we cannot include this observation for the test.
[25] Compared to previous research by Fehr and Gächter (2000) and Sefton et al. (forthcoming) the contribution levels in PuC are extremely high. This is especially noteworthy since it is has been shown in previous studies (e.g., Isaac and Walker 1988b) that cooperative behavior is much more difficult to establish if the community

clear increase in contributions in PuC whereas there is a decreasing trend ReC. The contributions in the second half increase significantly in PuC (Wilcoxon test, $p = 0.008$) whereas they decrease in ReC ($p = 0.031$). This finding supports the conjecture that the punishment possibility is a more effective mechanism for achieving cooperation in social dilemmas than the pure reward mechanism. However, the most striking result is that the members of PuC establish and maintain an almost perfect cooperation. In the second half, members of this community contribute on average 97.5% of their endowment. We observe no trend in the non-sanctioning communities. This is mainly due to sparse data since in the second half only very few subjects inhabit NSC in PUN.

Especially in PUN, there is a surprisingly clear differentiation in last round contributions of PuC and NSC. In the very last round of the experiment, members of PuC contribute virtually their complete endowment (100%)[26] whereas in NSC almost everything is kept in the private account (99%). Hence, we do not observe an endgame effect in PuC; in contrast, the contributions even rise. Without exception, all subjects in NSC in all treatments free ride ($g \leq 5$) in the last period. In ReC, the great majority (60.3%) free ride whereas there exist some subjects (36.5%) who also contribute high ($g \geq 15$). In NSC of REW, all subjects free-ride in the last period.

2.5.3. Payoffs and Efficiency

Figure 2.2 displays the development of the payoffs over the rounds for each treatment and for reasons of comparison it indicates the two benchmarks (social optimum and Nash equilibrium). An interesting observation is that in the first half of the experiment the payoffs in PuC are lower than in NSC (Wilcoxon test, $p = 0.148$). The higher initial contributions in PuC are "eaten up" by the higher expenses for punishment.[27] Hence, changing to PuC cannot be attributed to choosing the community that generates the highest average short-run payoff. Actually, we observe that in the first half of the experiment almost 40% of all changes from NSC to PuC are made although the last round average payoff in PuC is lower than in NSC. A change to the less profitable PuC cannot be explained by the prospect of higher future payoffs because subjects are free to change at any time in the game and they might also wait until the free riding diminishes. Joining PuC in the early stages with high contributions and high punishment expenses thus means the contribution to two joint projects with public good properties: to the joint project in Stage 1 of that period and to *cooperation enforcement* (second order public good) via punishment.

However, in the second half of the experiment the average payoff in PuC steadily increases towards the social optimum. In PuC, the payoffs in the second half of the experiment are significantly higher than in the first half (Wilcoxon test, $p = 0.008$). The payoffs in NSC remain constantly low from beginning and approach the Nash equilibrium in the end ($p = 0.219$). The payoffs in both communities of REW follow a clear decreasing trend. Payoffs fall significantly in ReC ($p = 0.031$). As in NSC of PUN, the decrease in NSC of REW is not significant ($p = 0.125$) since here the payoffs rapidly decrease already in the early phase of the experiment.

---

size is higher and the MPCR is low. Carpenter (2007), however, observes that cooperation is not necessarily reduced with an increasing group size if mutual monitoring and punishment possibilities are available.

[26] In that round, 88 out of 96 subjects who participated in PUN join PuC. 87 out of that 88 subjects contribute 20 in that round, the other one 19.

[27] This observation is in accordance to previous research (Sefton et al., forthcoming, Fehr and Gächter, 2002).

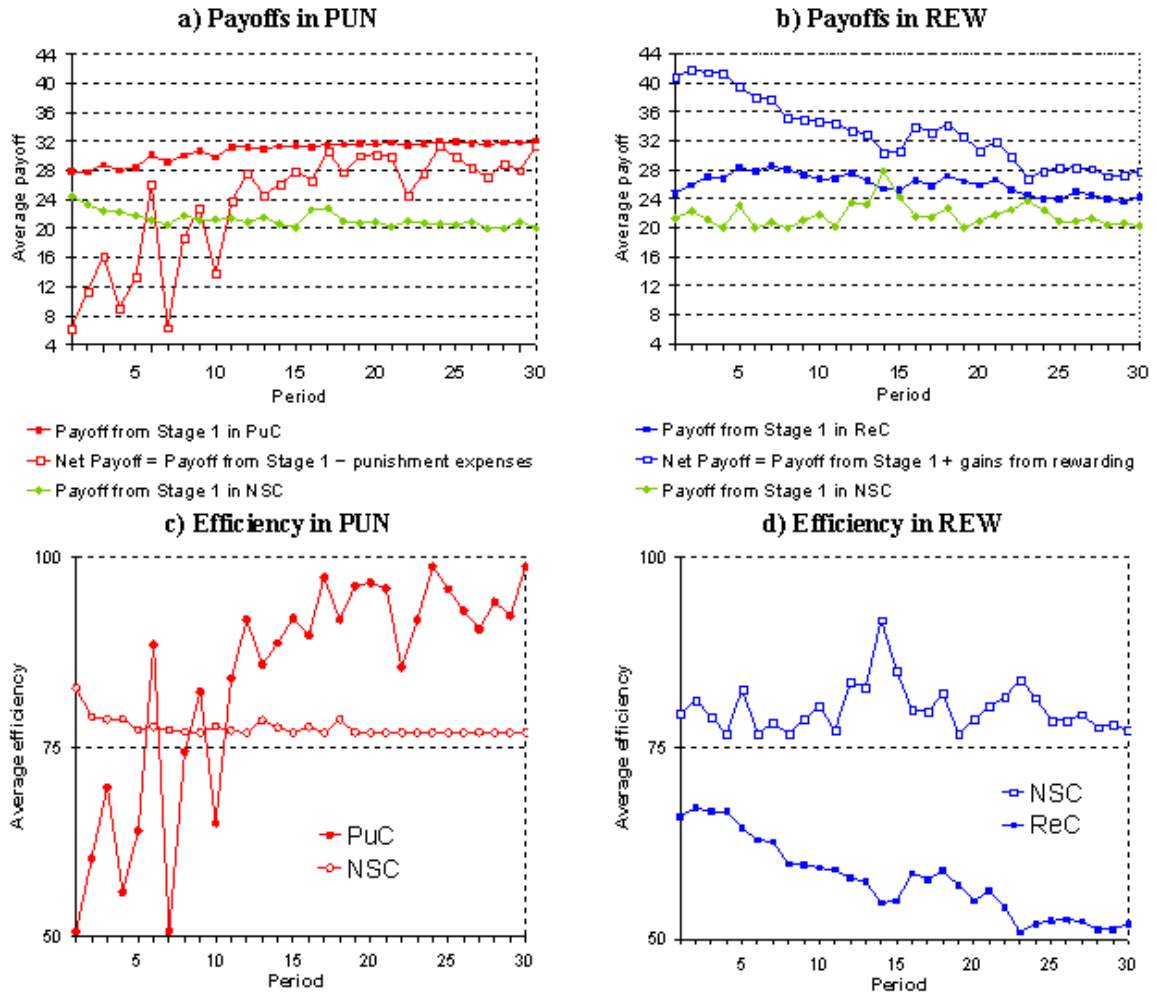


Figure 2.2 Payoffs and efficiency over periods

A social planner may be interested in the question which of the proposed treatments generates the highest social welfare. In the social optimum in PUN, each player fully contributes and refrains from punishing. This results in a payoff of 52 for each player, independent of the community size. In REW, however, rewarding is efficiency enhancing. Thus allocation of all rewarding tokens between the community members is socially optimal. In this case, the payoff of each player amounts to 92 tokens. Full contribution, however, does not constitute a Nash-equilibrium in either case. In the Nash equilibrium each player completely free rides and refrains from punishing and rewarding. This results in a payoff of 40 independent of the community and its size. We measure the *cooperative efficiency* (efficiency resulting from cooperation) as the ratio of the actually achieved surplus (Actual payoff – Nash Equilibrium) and the maximum possible surplus (Socially optimal payoff – Nash Equilibrium). In the second half of the experiment, in PUN, the cooperative efficiency amounts to 67.2% while it is only 17.7% in REW (U-Test, p = 0.003). Thus, the society offering the punishment community as an alternative performs best in the long run.

2.5.4. The comparison of theoretical predictions with experimental findings

To obtain further insights into the reasons why contribution behavior is so remarkably different in different communities, in the following we contrast the overall contribution behavior observed in the experiment to the predictions of the theoretical analysis. First, we investigate the PUN treatment.

Community choice in PUN

In the long-run, the most frequently observed situation in PUN is that a very small group of subjects inhabit NSC and defect while almost all subject are members of PuC and cooperate. This observation corresponds to the community choice equilibrium h) from Table 2.5 in which a defection equilibrium is played in NSC while a cooperation equilibrium is played with $g > 0$ in PuC. However, for one of the most frequent allocations observed in our experiment $n_S = 10$, the likelihood of equilibrium h) is below 0.01%. Thus, there is a striking difference between what we observe in PuC and what the theory predicts. In order to study this discrepancy more deeply, we will analyze the "evolution" of cooperation within PuC.

In the early phase of the experiment, i.e., in the first five periods, on average, 7.7 subjects choose NSC while 4.3 opt for PuC. If players' inequality parameters are distributed as in the adjusted FS distribution then for communities with $n_N = 8$ and $n_S = 4$ the theory predicts that cooperation is virtually never possible in NSC (likelihood below 1%) whereas in PuC the chances for cooperation are about 20%. Actually, however, in the early phase of the experiment, the cooperation *rate*, i.e., the percentage of endowment invested, amounts to 70.9% in PuC whereas it is 12.7% in NSC. In PuC, 45.6% of all contributions are full contributions (g = 20) and 68.6% are high contributions ($15 \leq g$). While 69.2% of contributions in NSC are low contributions ($g \leq 5$), 38.3% of them are even equal to zero. Thus, full contribution in PuC and zero contributions in NSC are the modal contributions. Hence, in the early phase, the experimental findings correspond to a certain extent to the theoretical predictions.

In PuC, the community size increases over the course of the rounds as Figure 2.1, panel a) shows. According to the theory, the probability for a cooperation-equilibrium in PuC should decrease when the community size increases. In the last five periods of the experiment, on average 11.2 subjects choose PuC. For such high occupations, the theory predicts virtually no cooperation (below 1%). However, the actual cooperation rate amounts to 98.4% in PuC. In the last period, we even observe perfect cooperation (all subjects contribute fully) in seven of the eight communities. In the eighth community; all but one subject contribute 20. How can this clear discrepancy between the experimental results and the theoretical predictions be explained?

According to the theory, a player j is ready to punish free-riders if j's "gains" from punishment outweigh the monetary costs of punishment c that j must bear (see condition (7) from Proposition 1). Thus, theoretically, for j it gets more and more difficult to satisfy the condition (7) the greater the number of the players in the community n_S gets. This means, a certain player j may satisfy the condition (7) for relatively small n_S, while the same player may not satisfy (7) when n_S increases. In the latter case, free-riders may not sufficiently threatened by punishment and thus have an incentive to deviate to lower contributions.

Figure 2.3, panel a) shows the percentage of players who punish *low* contributors (i.e., percentage of players who punish others who contribute lower) dependent on the community size. The theory predicts for the cost parameter used in our experiment ($c = 1/3$) that we should expect 10% or more punishers for small communities $n_S \leq 4$; for larger communities (as n_S grows) the chances for the existence of punishers diminish rapidly towards zero. However, in our experiment, we observe a completely different picture. For $n_S = 2$, the percentage of punishers in the experiment is much lower than predicted by the theory. This

could be due to the very rare occurrence of this occupation in our experiment. For $n_S \geq 3$, however, we observe that the percentage of *punishers* are much higher than predicted by the theory. Moreover, there is no decrease in the percentage of punishers when n_S increases. Actually, for $n_S \geq 3$, roughly 50% of subjects are punishers independent of n_S. Additionally, Figure 2.3, panel b) shows that the percentage of punishers remains relatively stable over time. Apparently, there are much more subjects who are ready to punish low contributors than predicted by the theory. How can we explain this?

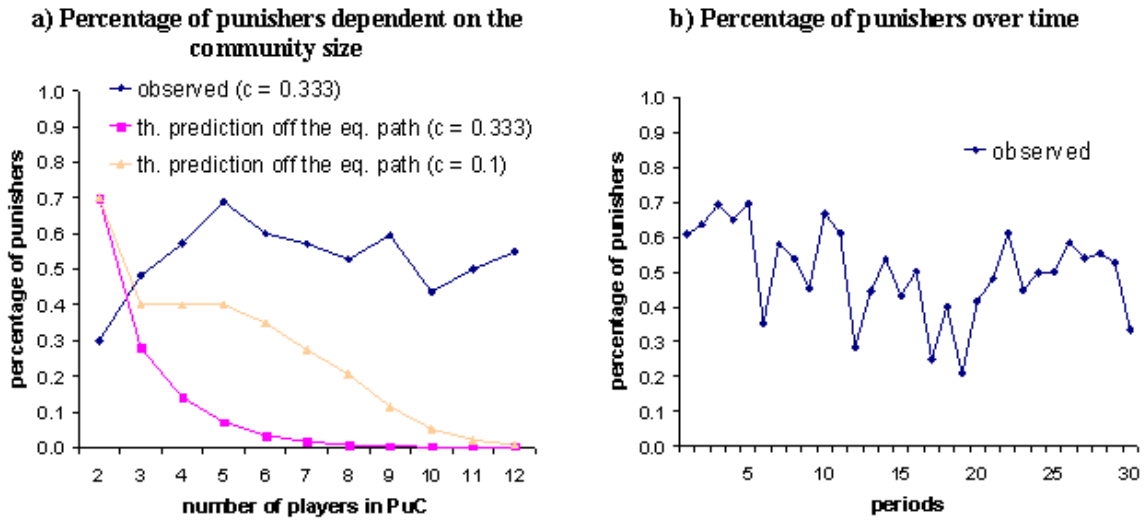


Figure 2.3 Percentage of subjects in PuC who punish low contributors

De Quervain et al. (2004) show that punishing norm violations may induce satisfaction for the punisher since punishment activates brain regions which are related to processing of "rewards". Moreover, subjects with a stronger activation are ready to spent more money on punishment. In light of this finding, one may hypothesize that some subjects may obtain additional utility from punishment that may be not appropriately reflected by the inequality aversion theory. Thus, for these subjects, the "actual costs" of punishing may be lower than the monetary costs of punishment. Can this be a source of explanation? Let us for a moment assume that there exist such subjects. How low must these actual costs be to have a considerable probability for a cooperation equilibrium? We set the punishment cost parameter extremely low $c = 1/10$ which makes punishing others extremely cheap. Figure 2.3, panel a) shows that this variation of c increases the theoretical predictions for the percentage of punishers. However, for large communities ($n_S \geq 10$), which we frequently observe in our experiment, the change in predictions are extremely small. Thus, the variation of c has virtually no effect on the theoretical probability for the existence of punishers and hence also on the probability for the existence of a cooperation equilibrium in PuC.

Do subjects possibly have inequality parameters higher than assumed in the FS distribution as well as in the adjusted FS and can these high parameters be "responsible" for the increased cooperation observed in the experiment? All other things held constant, a player j satisfies condition (7) the easier, the larger player j's α_j and β_j are. Let us assume that some subjects have larger α_j's and β_j's than in the adjusted FS distribution. Would this increase the theoretical probability for punishers in large communities? According to the FS model β_j must be lower than 1. To give cooperation the best chance let us assume that those players with the highest α_j and β_j combination shall have $\beta_j = 0.99$. According to the adjusted FS

distribution the fraction of players with the highest disadvantageous inequality parameter possess $\alpha_j = 4$. What happens if we increase their α_j? The right-hand side of the condition (7) converges to $1/(n_S - n_S')$ as α_j goes to infinity. Hence, for $\alpha_j \to \infty$ the condition (7) can be then stated as $c < 1/(n_S - n_S')$ which is equivalent to $n_S - n_S' < (1/c)$ for $n_S - n_S' > 0$. Thus, the effect of increasing α_j on the number of punishers is limited by the cost parameter. This means, the cost parameter determines the maximum number of punishers, hence also the number of free-riders in a community (see also Corollary 3). Hence, "adjusting" the inequality parameters does not considerably increase the percentage of punishers predicted by the theory.

To summarize: for communities that are frequently observed in PUN, the variations of *c* and the inequality parameters do not significantly increase the chances for the existence of the community-choice-equilibrium h). For $n_S = 10$, the probability for the existence of the equilibrium h) is never greater than 0.3%. Hence, the long-run allocations observed in PUN cannot be explained by the theory.

Community choice in REW

Now we contrast the theoretical predictions to the observed behavior in REW. In the long run, the most frequently observed situation in REW is defection in NSC and cooperation on a low level in ReC. This scenario corresponds to the community choice equilibrium h) from Table 2.6. Hence, theoretically, for one of the most frequently observed situations in our experiment, a community choice equilibrium with $n_S = 10$ in ReC has a probability of zero.

In the first five periods, on average, 10.7 subjects choose ReC while 1.3 subjects opt for NSC. For $n_S = 10$, according to the theory, chances for a cooperation equilibrium in ReC amount to 4.7% . The actual cooperation rate averaged over the first five periods is 55.1% in ReC whereas 24.1% in NSC. Thus, in the beginning, in contrast to the theoretical predictions there is a considerable amount of cooperation in ReC.

Figure 2.4, panel a) shows, dependent from the community size, the percentage of subjects who should reward others according to the theory (on the equilibrium path) and the actual percentage of reward providers in the experiment (here we count subjects who reward a fellow community member who contribute exactly the same amount or more). For large communities as frequently observed in the experiment, the theory predicts that a very small fraction of players would reward (below 5%). However, in the experiment, we actually observe that 40%-50% are reward providers.

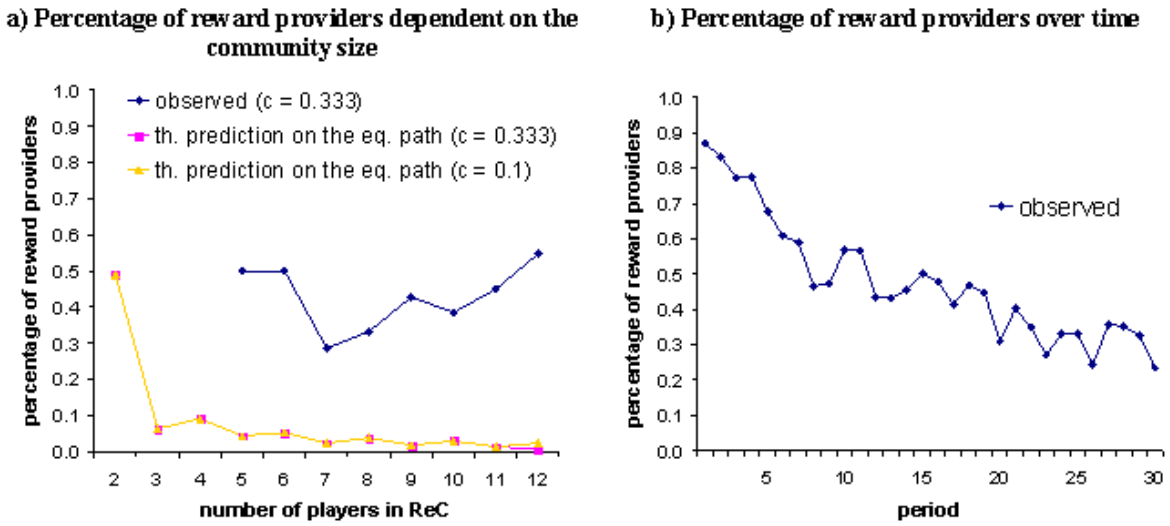


Figure 2.4 Percentage of subjects in ReC who reward others with equal or more contributions

Unlike PuC, the percentage of subjects who opt for ReC is high in the beginning and remains stable over the rounds. In the last period, on average, 10.5 subjects inhabit ReC while 1.5 subjects are members of NSC. Thus the occupation of ReC is roughly the same as in the early phase. However, over the rounds, we observe a strong decline in contributions in ReC. In the last five periods, the cooperation rate in ReC amounts to 35.3% while it is 8.6% in NSC. Figure 2.4, panel b) shows that the decline in contributions is accompanied by a strong decrease of the percentage of reward providers. While in the first five periods 78.5% of subjects are reward providers in the last five periods only 30.3% reward. As Figure 2.4, panel a) shows, the variation of the cost parameter for rewarding has virtually no effect on the percentage of reward providers.

For $n_S = 10$, the theory predicts that the probability for the existence of the equilibrium h) is virtually zero. Thus, in REW (as in case of PUN), in the long run the theory and the experimental findings diverge. In REW, as in PUN, variations of c and α_j does not significantly change the probabilities for the community choice equilibrium h).

Summary of the comparison of the theory with the observed community choice

To sum up, in PUN as well as in REW, in the long run the experimental findings are different from the theoretical predictions. In both treatments, "adjusting" the parameters of the model does also not help to bridge the gap between the predictions of the theory and the observed behavior in our experiment. In the following, we deepen the analysis of subjects' behavior by investigating the immediate changes in contributions and sanctioning after a community change.

2.5.5. Community changes and immediate change in contributions

In PuC, in the early phase of the experiment, a "culture of cooperation" is established by enforcers who harshly punish free-riders. Already in the first period, the vast majority (87.5%), of the high contributors in PuC exert punishment tokens to discipline low contributors. These subjects amount to 14.6% of the total population. The punishment of defectors is likely to serve two purposes: it either encourages the defectors to increase their contributions or it convinces them to leave PuC. Subjects receive the more punishment tokens

the more they deviate from the respective community average.[28] Thus on average, any attempt to free ride is not tolerated by the other community members. In subsequent periods, a majority of the incoming subjects to PuC adapt this "culture", i.e., they not only contribute high but also participate in the second order public good and punish free-riders.

Figure 2 panel a) illustrates the cumulative distribution of the changes in contributions immediately after a subject has moved to the other community in PUN.[29] In total 76.3% of subjects increase their contribution after switching from NSC to PuC. 46.4% of the subjects contribute very low ($g \leq 5$) in NSC before switching; but the same subjects contribute very high ($g \geq 15$) after the change in PuC. Moreover, 17.9% even "convert" from a complete free-rider to a full cooperator; increasing their contributions from 0 to 20 after the change to PuC. In 46.7% of these cases, subjects punish other subjects who contributed less than they themselves immediately after change. Thus, the former free-riders not only adapt their contributions but they also engage in costly punishment of subjects whose payoff from Stage 1 is higher than the own payoff, i.e., they indeed reduce the inequality in payoffs between themselves and others in the FS sense. The change in contributions after switching in the opposite direction, i.e., from PuC to NSC, is similarly extreme: 66.8% of subjects decrease their contribution after the change to NSC. 15.3% switch even from full cooperation to complete free-riding.[30]

The predominant tendency to punish free-riders after a migration from NSC to PuC nicely demonstrates that subjects adapt to the common behavior in PuC. This "conformist" tendency sustains punishment activities on a considerable level such that cooperation can be stabilized. As the size of PuC grows, there exist always a sufficient number of subjects who are ready to punish the free-riders.[31]

In REW, a change between communities barely has an effect on subjects' contributions (see Figure 2 panel b). After switching to ReC, only 39.2% of subjects increase their contributions at all. When the change is directed from ReC to NSC, roughly one third of the subjects increase their contributions while one third decrease their contributions after the change. Another third do not change their investments at all.

---

[28] All Spearman-rank-correlation-coefficients computed for each independent observation are positive. A binomial test with these correlation coefficients rejects that a negative correlation is as likely as a positive correlation ($p = 0.016$). Below average contributors are punished with 10.4 tokens. Moreover, the subjects who contribute less than 15 tokens compared to the community average are punished with more than 30 punishment tokens on average, i.e., they have to suffer from an income reduction of more than 90 tokens. This means that they incur a true loss in that round since the saved contribution of 20 token is by far out weighted by the received punishment. However, above average contributors also receive 0.7 tokens, on average. For the frequently observed punishment of high contributors in experiments see e.g. Cinyabuguma et al. (2006).

[29] We calculate the difference between the subject's contribution in round t and the contribution of that subject in round t -1 conditional upon the subject changed the community from round t -1 to round t.

[30] This observation is well in line with the experimental findings by Falk et al. (2003). They show that the contribution behavior of subjects is influenced by social interactions with their "neighbors". Subjects who simultaneously belong to two different groups with disjoint group compositions exhibit conditionally cooperative behavior, i.e. the same subject contributes more if she is in a community with high contributors and contributes less if she is in a community with low contributors.

[31] The burden of punishing is shared: 92 out of 95 subjects who inhabit PuC faced at least one time a situation in which a community member contributed less than themselves. In such a situation, 81.5% of them punished at least one time a low contributor, 69.6% punished more than once in such a situation. More than half of the subjects (55.4%) even punished 5 or more times.

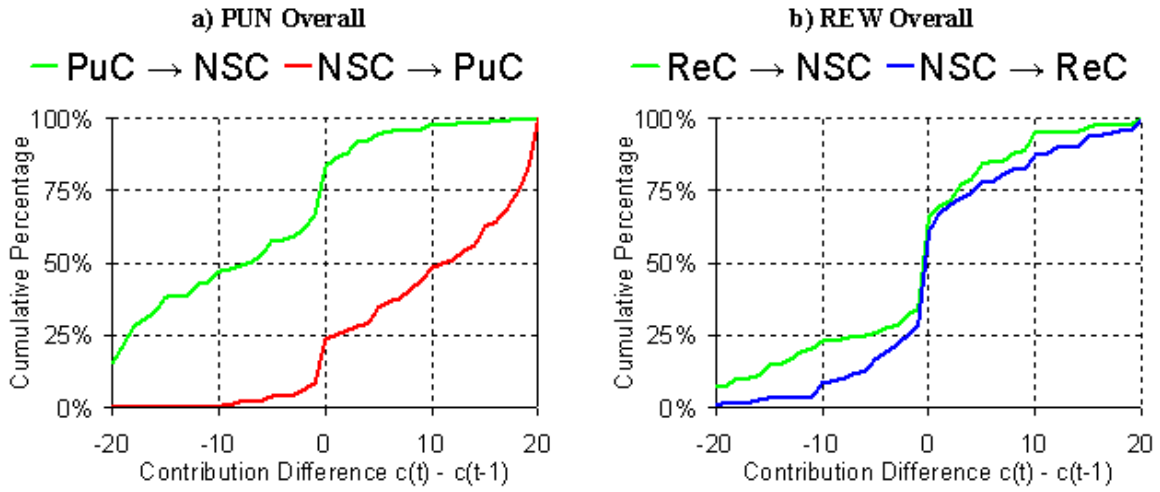


Figure 2.5 Cumulative distribution of contribution differences after the community change

The way how altruistic rewarding might enhance cooperation is different. While punishment is mainly addressed to free-riders who act in their own but not in the groups' interest, rewarding is directed to group members who already act to the best of the group. In 66.1% of all cases, high contributors reward other high contributors. The vast majority of rewards are directed to subjects who contribute equal or more than the reward provider (in 58.1% and 30.2% of all cases, respectively). If low contributors are rewarded at all, they receive very small amounts. Below average contributors receive 0.7 tokens whereas above average contributors receive 6.1 tokens. Below average contributors send on average 1.8 reward tokens to other community members whereas above average contributors send 5.0 tokens. Hence, rewarding is mainly an efficiency enhancing exchange of transfers within the group of cooperative subjects. This fact constitutes a fundamental difference between the use of punishment and rewards. While punishment is addressed to norm violators in order to "encourage" them to cooperate, rewarding is addressed to the subjects who already cooperate. Non-cooperators may only be indirectly convinced to cooperate by experiencing and/or observing that cooperators receive additional rewards. This indirect mechanism does not seem to be as effective as the direct mechanism of punishment. In fact, interval regressions (see Table 2.12) with observations as clusters show that punished subjects increase their contributions in the following period whereas rewarded subjects decrease them.

Table 2.12: Contribution difference of the teammate between t–1 and t

Coefficient	PuC	ReC
Period	-0.009 (0.006)	-0.126*** (0.026)
Received tokens in *t*-1	0.339*** (0.064)	-0.393** (0.171)
Constant	0.104 (0.125)	3.489*** (0.834)

***, **, and * denote significance level 0.01, 0.05, and 0.1 respectively.
Robust standard errors in parentheses (adjusted for 8 clusters in PuC and for 6 clusters in ReC).

2.6. Conclusion

In this study we theoretically and experimentally investigate the community choice in a social dilemma setting. Our theoretical considerations predict that equilibria with positive contributions occur under very restrictive conditions. In particular, cooperation in large communities is virtually never possible given the commonly accepted distributions of inequality parameters in the population. The reason is that in general, to sustain cooperation in large communities, a great fraction of subjects must have high inequality parameters. Our

experimental findings, however, show a different picture. In PUN – where subjects are given the choice between a non-sanctioning community (NSC) and a community with punishment possibilities (PuC) – initially, most subjects choose NSC, but over time PuC becomes selected with an increasing frequency. In the final rounds almost all subjects join PuC. Despite severe punishment in the beginning, PuC leads to high efficiency levels with full contribution of all participants and no punishment in the end. In REW, where the choice is between NSC and a community with rewarding possibilities (ReC), we do not observe a comparable effect. Despite subjects' strong initial preference for ReC and a relatively high level of cooperation – maintained by a frequent rewarding activity – over time subjects' willingness to reward deteriorates. This is surprising since rewarding is efficiency enhancing. We observe that on average exchanging rewards does not lead to an increase in contributions but it rather activates a spiral downwards which leads over time to a considerable decrease of cooperation.

Models of cultural group selection indeed show that sanctioning defectors can result in stable cooperation when strategies for cooperation and punishment can spread among individuals through social learning (see e.g., Boyd and Richerson 1992 or Henrich and Boyd 2001). If the norm to cooperate and punish defectors is sufficiently manifested in the group culture and punishment is sufficiently costly for the punished subject, rare non-cooperators are less successful than cooperators because they experience severe punishment. A tendency to imitate successful behavior (payoff-biased transmission) enhances the propensity of cooperative behavior. Punishment behavior is stabilized by the tendency to imitate the common behavior (conformist transmission). In the latter case group members punish, because it is common to do so, although cooperators who do not punish defectors have a payoff advantage compared to those who both cooperate and punish. However, when cooperation in a group is sufficiently widespread, this payoff-advantage is relatively small and only a weak tendency of conformist behavior suffices to oppose the payoff-biased transmission (in small disfavor of punishers) and stabilize the punishment of non-cooperators. Cooperation and punishment, however, does not constitute the only equilibrium in cultural group selection; non-cooperation and non-punishment can also be sustained as an equilibrium.

Our experimental findings from the PUN treatment seem to support the intuition provided by the payoff-biased and conformist transmission. Cooperative groups start with few members and are then invaded by free-riders who are attracted by high payoffs. Once being a member of PuC, subjects seem to change their types – although they used to refrain from contributing in NSC, once being in PuC they convert to a cooperator and even into a punisher. Since more and more subjects share the burden of punishment, the payoff disadvantage of punishers diminishes. The consequence is that we observe almost full cooperation in a highly occupied punishment community.

With this study we provide a further step towards the understanding of community choice in social dilemmas. Our results provide evidence that punishment mechanism "outcompetes" the non-sanctioning mechanism when both can be selected. In an indirect way of comparison, punishment is also more successful than a community which relies on voluntary bilateral rewarding because it leads to a very high level of cooperation. The latter does better if it can be selected against a non-sanctioning mechanism, however by far not as good as the punishment mechanism. This result expands previous findings on the superior performance of punishment mechanisms when they are exogenously "imposed".

Our results raise several promising questions for future research. How sensitive are our findings regarding the switching costs between communities? How important is the role amount of information about the other community? Or, what happens if individuals cannot choose between communities, but instead can create new alternative institutions or adapt the

existing institutional framework? How can the initial losses in PuC be mitigated? Contributions in these directions can help us understand how institutions emerge, work, and how they influence the human behavior and pave the way for designing effective institutions that reduce the downside effects inherent to social dilemmas.

2.7. References

Abbink, Klaus; Irlenbusch, Bernd, and Renner, Elke. "The Moonlighting Game – An Experimental Study of Reciprocity and Retribution." *Journal of Economic Behavior and Organization*, 2000, 42, pp. 265-277.

Ahn, T.K., Isaac, M., Salmon, T. "Endogenous Group Formation" Forthcoming in *Journal of Public Economic Theory*.

Andreoni, James; Harbaugh, William, and Vesterlund, Lise. "The Carrot or the Stick: Rewards, Punishment, and Cooperation." American Economic Review, 2003, 93(3), pp. 893-902.

Bellemare, C., Kröger, S., and, van Soest, A. "Measuring Inequity Aversion in a Heterogeneous Population using Experimental Decisions and Subjective Probabilities" Working Paper, 2007.

Berg, J., Dickhaut, J., and McCabe, K. "Trust, Reciprocity and Social History." *Games and Economic Behavior*, 1995, *10*(1), pp. 122-42.

Blanco, M., Engelmann, D., and Normann, H-T. "A Within-Subject Analysis of Other-Regarding Preferences" Working Paper, 2007, Royal Holloway, University of London.

Botelho, A., Harrison, G., Costa Pinto, L.M., and Rutström, E.E. "Social Norms and Social Choice." Working Paper, 2007.

Boyd, R. and Richerson, P. J. "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups." *Ethology and Sociobiology*, 1992, *13*(3), pp. 171-95.

Brandts, J. and Schram, A. "Cooperation and noise in public goods experiments: applying the contribution function approach." *Journal of Public Economics*, 2001, *79*, pp. 399-427.

Brosig, J., Ockenfels, A., and Weimann, J. "The Effect of Communication Media on Cooperation." *German Economic Review*, 2002, 4, pp. 217-241.

Brown, M., Falk, A., and Fehr, E. "Relational Contracts and the Nature of Market Interactions." *Econometrica*, 2004, *72*, pp. 747-780.

Camerer, C. "Behavioral Game Theory: Experiments on Strategic Interaction." Princeton: Princeton University Press, 2003.

Carpenter, J. "Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods", *Games and Economic Behavior*, 2007, 60(1), pp. 31-52 (2007).

Charness, G., Yang, C-L., "Endogenous Group Formation and Public Goods Provision: Exclusion, Exit, Mergers, and Redemption", Working Paper, 2007.

Cinyabuguma, M., T. Page and L. Putterman. "Can second-order punishment deter perverse punishment?" *Experimental Economics*, 2006, V9(3): pp. 265.

Coricelli, G., Fehr, D., and Fellner, G. "Partner Selection in Public Goods Experiments." *Journal of Conflict Resolution*, 2004, *48*(3), pp. 356-378.

Croson, R.T.A. "Partners and strangers revisited." *Economics Letters*, 1996, *53*, pp. 25-32.

Croson, R.T.A. "Theories of Altruism and Reciprocity: Evidence from Linear Public Goods Games." *Working Paper*, 1998, *98-11-04*, The Wharton School of the University of Pennsylvania.

de Quervain, D. J. F., Fischbacher, U., Treyer, V., Schelthammer, M., Schnyder, U., Buck, A., and Fehr, E. "The Neural Basis of Altruistic Punishment." *Science*, 2004, 305(5688), pp. 1254-58.

Dawes, R.M. “Social Dilemmas.” *Annual Review of Psychology*, 1980, *5*, pp. 163-193.

Ehrhart, K-M., and Keser, C. “Mobility and Cooperation: On the Run.” *Working Paper 99s-24*, 1999, CIRANO, Montreal.

Ertan, A., Page, T., Putterman, L. “Can Endogenously Chosen Institutions Mitigate the Free-Rider Problem and Reduce Perverse Punishment?” 2005, Working Paper 2005-13, Department of Economics, Brown University.

Falk, A., Gächter, S., and Fischbacher, U. “Living in Two Neighborhoods - Social Interactions in the Lab.” *Working Paper*, 2003, *No. 150*, University of Zurich.

Fehr, E., and Schmidt, K. “A theory of fairness, competition, and cooperation.” *Quarterly Journal of Economics*, 1999, *114*(3), pp. 817-868.

Fehr, E., Fischbacher, U., and Gächter, S. “Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms.” *Human Nature-an Interdisciplinary Biosocial Perspective*, 2002, *13*(1), pp. 1-25.

Fehr, E., and Gächter, S. “Cooperation and Punishment in Public Goods Experiments.” *American Economic Review*, 2000, *90*(4), pp. 980-994.

Fehr, E., and Gächter, S. “Altruistic Punishment in Humans.” *Nature*, 2002, *415*, pp. 137-140.

Fehr, E., Kirchsteiger, G., and Riedl, A. “Gift Exchange and Reciprocity in Competitive Experimental Markets.” *European Economic Review*, 1998, *42*(1), pp. 1-34.

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

Gächter, Simon and Fehr, Ernst. “Collective action as social exchange.” *Journal of Economic Behavior and Organization*, 1999, *39*, pp. 341-369.

Gürerk, Ö., Irlenbusch, B. and Rockenbach, B. “The Competitive Advantage of Sanctioning Institutions.” *Science*, 2006, 312(5770): pp. 111.

Hardin, G. “The Tragedy of the Commons.” *Science*, 1968, *162*, pp. 1243-1248.

Hauk, E. and Nagel, R. “Choice of Partners in Multiple Two-Person Prisoner’s Dilemma Games.” *Journal of Conflict Resolution*, 2001, *45*(6), pp. 770-793.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., and McElreath, R. “In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies.” *American Economic Review*, 2001, *91*(2), pp. 73-78.

Henrich, J., and Boyd, R. “Why People Punish Defectors - Weak Conformist Transmission Can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas.” *Journal of Theoretical Biology*, 2001, *208*(1), pp. 79-89.

Isaac, M., and Walker, J. “Communication and Free Riding Behavior: The Voluntary Contributions Mechanism.” *Economic Inquiry*, 1988a, *26*(2), pp. 585-608.

Isaac, M., and Walker, J. “Group Size Hypotheses of Public Goods Provision: An Experimental Examination.” *Quarterly Journal of Economics*, 1988b, *103*, pp. 179-199.

Keser, C., and van Winden, F. “Conditional Cooperation and Voluntary Contributions to Public Goods.” *Scandinavian Journal of Economics*, 2000, *102* (1), pp. 23-39.

Kirchsteiger, G., Niederle, M., and Potters, J. “Endogenizing Market Institutions: An Experimental Approach).” *European Economic Review*, Vol 49(7), October 2005, 1827-1852.

Ledyard, J. “Public Goods: A Survey of Experimental Research.” J. Kagel and A. Roth, *Handbook of Experimental Economics*, 1995, Princeton University Press, pp. 111-194.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M-C. "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review*, 2003, *93*(1), pp. 366-380.

Orbell, J.M., van de Kragt, A.J.C., and Dawes, R.M. "Explaining Discussion Induced Cooperation." *Journal of Personality and Social Psychology*, 1988, *54*(5), pp. 811-819.

Ostrom, E. "A Behavioral Approach to the Rational Choice Theory of Collective Action." *American Political Science Review*, 1998, *92*(1), pp. 1-23.

Ostrom, E., Walker, J., and Gardner, R. "Covenants with and without a Sword: Self-Governance is Possible." *American Political Science Review*, 1992, *86*, pp. 404-417.

Page, T., Putterman, L., and Unel, B. "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency." *The Economic Journal*, 2005, 115 (October), pp. 1032–1053.

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

Riedl, A. and Ule, A. "Exclusion and Cooperation in Social Network Experiments." mimeo, 2003, University of Amsterdam.

Sefton, M., Shupp, R., and Walker, J. "The Effect of Rewards and Sanctions in the Provision of Public Goods." Forthcoming in *Economic Inquiry*.

Sutter, M., Haigner, S., and Kocher, M.G. "Choosing the carrot or the stick? – Endogenous institutional choice in social dilemma situations." CEPR Discussion Paper, 2006.

Tiebout, C.M. "A Pure Theory of Local Expenditures." *Journal of Political Economy*, 1956, *64*, (5), pp. 416-424.

Weimann, J. "Individual behaviour in a free-riding experiment." *Journal of Public Economics*, 1994, *54*, pp.185-200.

Yamagishi, T. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology*, 1986, *51*(1), pp. 110-116.

3. Social history and the community choice in social dilemmas

3.1. Introduction

For communities with an exogenously fixed composition of members, experimental studies identify peer punishment as valuable means to sustain cooperation in social dilemma situations (Yamagishi 1986, Ostrom et al. 1992, Fehr and Gächter, 2000). Recent studies by Gürerk et al. 2006 and 2007 show that communities with punishment possibilities even prevail in societies with competing communities applying different institutional rules. In their experimental setup, subjects repeatedly choose *between* communities with and without punishment possibilities before interacting in a social dilemma situation. In the beginning, the great majority of individuals prefer the non-punishment community (NPC). However, over time the situation changes dramatically. Since cooperators who also bear costs to punish free-riders cause cooperation to rise in the punishment community (PuC), more and more individuals leave NPC and join PuC. Towards the end of the experiment, more than 90% of subjects inhabit PuC. Even though the community becomes very large, the cooperation rate remains surprisingly stable and converges to almost 100%. This finding clearly demonstrates the competitive advantage of peer punishment communities. However, cooperation enforcement through punishment is costly and lowers efficiency in the short run. The question arises how this problem can be mitigated.

There are several possible explanations for subjects' initial reluctance to join the punishment community. Subjects may have a "natural aversion" against receiving punishment since they associate negative feelings with it. Social psychologists define negative sanctions as deliberate acts that lead to unpleasant inner states that the punished person wants to avoid. Additionally, subjects may fear to be exposed to "perverse" punishment, i.e., punishment of high contributors. In fact, this is a frequently observed phenomenon in experimental studies (see e.g. Cinyabuguma et al. 2006).

On the other hand, subjects may anticipate that there will be a lot of punishment in PuC, especially in the beginning of the interaction. For this reason, they may "rationally" decide to wait and see how the situation in PuC develops before joining in. There is no reason to wait anymore when the payoffs in PuC outperform the payoffs in NPC. From that point it is more profitable to join PuC than to remain in NPC. Yet another explanation for the initial reluctance could be that subjects simply do not anticipate correctly that PuC is the more efficient community in the long run. Previous studies show that the provision of a social history about the results of an experiment may in some situations influence the behaviour of the informed subjects who are going to play the same game (Berg et al. 1995), whereas it does not influence the results in others (Fehr and Rockenbach 2003).

In this study, we investigate whether the lack of information on the high cooperation levels in PuC is a critical force that drives the reluctance to join PuC. Our conjecture is that social history, i.e., experience-based information, will lower people's reluctance towards PuC by correcting the possible "false anticipation" about its performance. In order to test this we conduct a social history treatment (SH) which is the exact replication of the PUN treatment from the study Gürerk et al. 2007. The only difference is that in SH a social history – which contains the main results of the PUN – is provided to the subjects.

3.2. The Basic Model and the Experimental Design

The experiment is based on a public goods game of 30 repetitions including three stages in each period: In Stage 1 (institution choice stage) N participants in each group choose (without a cost) between a non-punishment community (NPC) and a community with a peer punishment mechanism (PuC). In Stage 2 (contribution stage) members of each community are endowed with 20 experimental tokens and can anonymously invest g ($0 \le g \le 20$) in the public good. The defining characteristic of a public good is fulfilled independently from the number of members n^{θ} with $\theta \in \{1,2\}$ in each community because the marginal per capita return a is $1/n_i^{\theta} < a < 1$ for all n_i^{θ} with $2 \le n_i^{\theta} \le N$. In PuC, Stage 2 is followed by Stage 3 (punishment stage). Here, all subjects are endowed with 20 additional tokens and may anonymously assign punishment tokens to each other (subjects in NPC also receive additional 20 tokens and simply keep these). Each received punishment token lowers the payoff of the punished subject by three tokens. After each period, all participants receive feedback about contributions, received punishment tokens and payoffs in *both* communities. For a money maximizing subject with self-centered preferences it is always the dominant strategy to refrain from contribution as well as not to punish.

The social history is handed out to subjects before the experiment starts. For both communities it separately tabulates the averages of the number of community members, contributions, received punishment tokens in PuC, and the payoffs of the baseline treatment (PUN) for each period. Additionally, the developments of the averages are visualized in small figures. In the instructions we use a neutral framing by labeling the communities A and B and we refrain from using the word punishment. 72 students from University of Erfurt participated in three experimental sessions. A session lasted on average 2 hours including instructions. Average earnings were 24 Euros.

3.3. Results

Already in the first period the social history has significant effects on subjects' community choices and contributions. The majority (54.2%) of the participants in SH prefer PuC to NSC in the first period. This is in sharp contrast to the baseline treatment PUN where only less than one third (31.2%) of the subjects opt for PuC ($p = 0.037$)[32]. Thus, social history lowers the initial reluctance against PuC significantly. Moreover, subjects in PuC of SH cooperate on a higher level than subjects in PuC of PUN in the first period. In particular, the fraction of high contributions ($g \ge 15$) is significantly larger in SH (74.4%) than in PUN (53.3%) ($p = 0.057$). On the other hand, in SH only 5.1% of the subjects who join PuC initially are free-riders ($g \le 5$) whereas in PUN this is true for one third of the subjects. Thus, when subjects are provided with a social history, PuC attracts cooperators while it deters free-riders. In total, this results in a very high cooperation level of 78.5% in PuC.

Let us now take a closer look at the initial punishment behavior. Does the social history lead to less punishment (because of the high cooperation level that is achieved immediately) or does it lead even to more punishment (because the social history encourages the members of PuC to punish)? Indeed, the frequency of peer-to-peer punishment is slightly higher in SH (2.1 instances per subject) than in PUN (1.4) in the first period. However, the severity of punishment, i.e., average tokens sent per punishment instance, is lower in SH (2.3 tokens) than in PUN (3.4) ($p = 0.131$). Hence the amount of tokens spent for punishment per member is roughly the same in both treatments (5.1 in SH, 5.4 in PUN).

[32] All reported non-parametric statistical tests are two-tailed Mann-Whitney U-tests and use the session averages as independent observations.

In both treatments, the average payoff in PuC (29.2) is significantly lower than in NPC (44.9) ($p = 0.031$) in the first period. Although this payoff difference between PuC and NPC is less pronounced in SH, subjects in both treatments suffer from a considerable efficiency loss due to punishment. However, in SH, average earnings in PuC rapidly catch up the payoffs in NPC. Already in the period 5, members of PuC earn on average more than members of NPC (see Figure 3.1d). Moreover, in SH, from period five onwards the average payoffs in NPC are constantly lower than in PuC. In contrast, in PUN, the average payoffs in PuC oscillate strongly in the beginning. They catch up with the payoffs in NPC in period 11 (see Figure 3.1c). The average period where the earnings in PuC exceed the earnings in NPC is 7.2 in SH, while it is 15.1 in PUN ($p = 0.037$). This relation between earnings continues for the rest of the experiment. Hence, PuC becomes the more profitable community much earlier in SH than in PUN. Since PuC becomes the more profitable community, more and more subjects migrate from NPC to PuC in both treatments (see Figure 3.1b).

With exception of one group in each treatment, there is at least one period in which all group members join PuC. However, such a period is observed much earlier in SH (on average in period 9.6) than in PUN (17.6) ($p = 0.033$). Moreover, the average number of consecutive periods where all members of a group inhabit PuC amounts to 15.4 periods in SH but only to 5.6 periods in PUN ($p = 0.027$). In the last period, only one single subject in SH is not a member of PuC whereas in PUN roughly 10% of subjects still inhabit NPC.

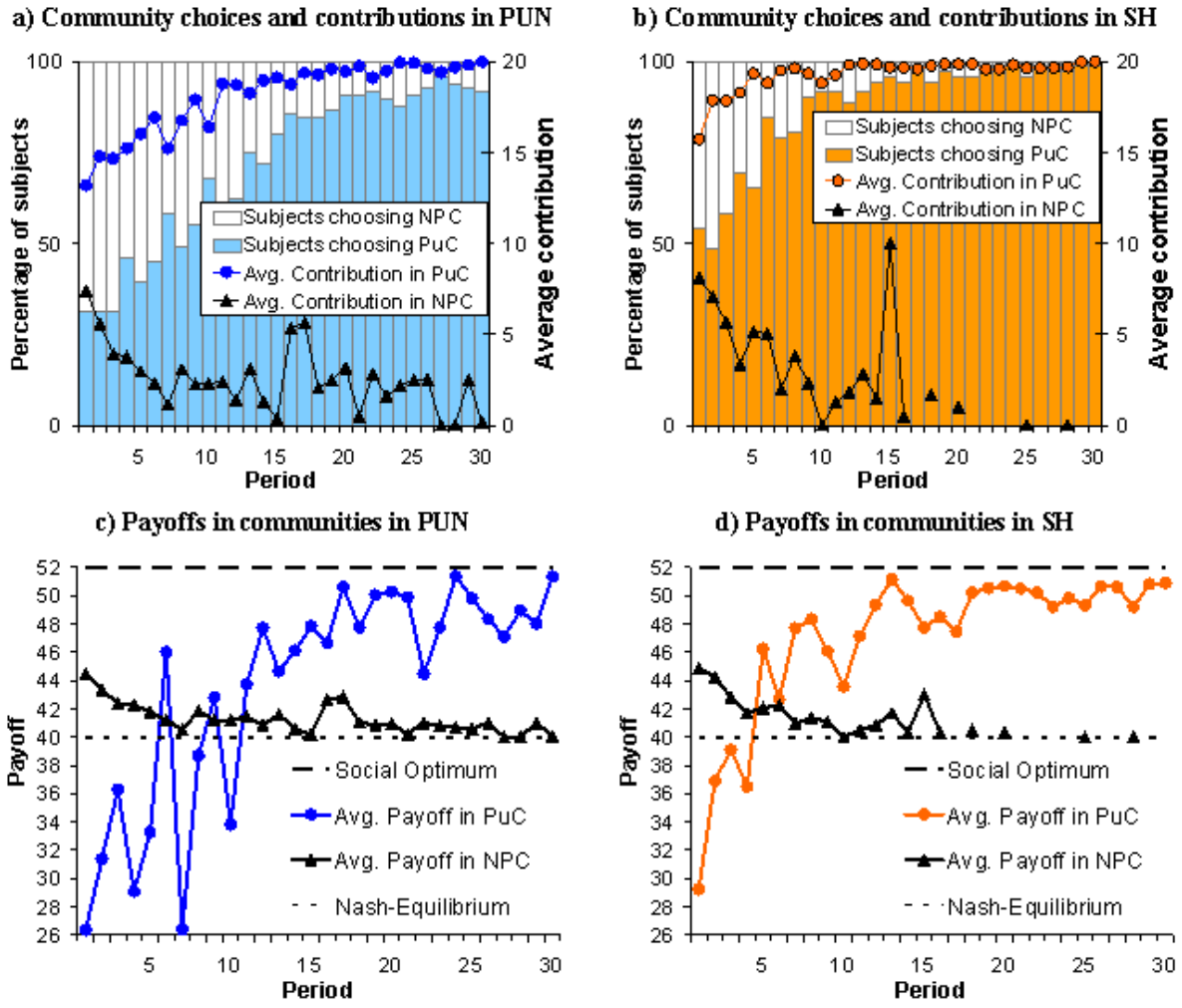


Figure 3.1 Average numbers

If we look at total averages, punishment in SH is less severe than in PUN. A punished person receives less tokens in SH (5.2) than in PUN (6.8). Moreover, the severity of a punishment

instance is significantly lower in SH (2.4 tokens) than in PUN (3.1) ($p = 0.043$). In both treatments, subjects who contribute more frequently punish subjects who contribute less. In SH, this is true for 38.5% of relative instances while it is only true in 25.5% in PUN ($p = 0.101$). Interestingly, in SH subjects with very high contributions are punished. On average a subject who receive punishment tokens contributes more in SH than in PUN ($p = 0.064$). The frequency of "perverse" punishment, i.e., the cases when the punished person contributed equally or more than the punisher is roughly the same in both treatments (SH: 1.7%, PUN: 1.8%). However, the average severity of the unjustified punishment instances is significantly lower in SH (1.5 tokens) than in PUN (1.9) ($p = 0.075$).

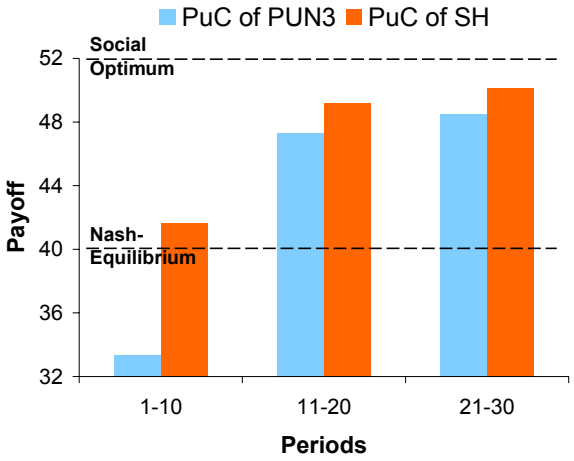


Figure 3.2 Overall payoffs

From the social planner's perspective, one of the most interesting issues concerns the efficiency in the "society" as a whole including all its communities. Figure 3.2 shows the evolution of groups' overall earnings in both treatments. In SH, the average payoff amounts to 47.1 tokens while it is 44.5 in PUN. The surplus ratio, i.e., the actual surplus generated by cooperative behavior in the experiment divided by the maximum possible surplus amounts to 59.2% in SH and 37.5% in PUN. Thus, the gains from cooperation are clearly higher in SH than in PUN ($p = 0.101$).

3.4. Conclusions

In this study, we explore whether informed subjects are less reluctant to join a community with a peer punishment mechanism in a social dilemma situation than uninformed subjects. We find that significantly more informed subjects join the punishment community initially and start to cooperate on a much higher level than the uninformed. With social history, the punishment is more directed at low contributors and is not as severe as in the baseline treatment. This results in higher payoffs. In the long run, with social history, punishment community attracts even more members of the society compared to when social history is not provided. As a consequence, the societies reach a higher overall efficiency. Our findings shed light on the importance of experience-based information for the acceptance of seemingly unpopular but socially desirable mechanisms.

3.5. References

Berg, J., Dickhaut, J., and McCabe, K. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 1995, 10(1): pp. 142.

Cinyabuguma, M., Page, T., and Putterman, L. "Can second-order punishment deter perverse punishment?" *Experimental Economics*, 2006, V9(3): pp. 265.

Fehr, E. and Gächter, S. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 2000, 90(4): pp. 994.

Fehr, E. and Rockenbach, B. "Detrimental effects of sanctions on human altruism." *Nature*, 2003, 422(6928): pp. 140.

Gürerk, Ö., Irlenbusch, B., and Rockenbach, B. "The Competitive Advantage of Sanctioning Institutions." *Science*, 2006, 312(5770): pp. 111.

Gürerk, Ö., Irlenbusch, B. and Rockenbach, B. "On community choice in social dilemmas – A voting with feet approach." Working Paper, 2007, University of Erfurt

Ostrom, E., Walker, J. and Gardner, R. "Covenants With and Without a Sword: Self-Governance is Possible." *American Political Science Review* 1992, 86(2): pp. 417.

Yamagishi, T. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology*, 1986, 51(1): pp. 110-116.

4. Motivating Teammates: The Leader's Choice of Positive and Negative Incentives[33]

"*The […] most obvious way to bring about cooperation between employees […] is to pay for cooperation or to punish uncooperative behavior.*"

Edward P. Lazear[34]

4.1. Introduction

In his classic study of leadership behavior Burns (1978) combines insights from the literature on traits, leadership styles and leader-member exchange research and coins the distinction between *transactional* and *transformational* leadership styles. In Burn's view these two styles constitute different poles of leadership style dimensions. A transformational leader is one who offers a purpose that transcends short-term goals and focuses on higher order intrinsic needs. Exhibiting transactional leadership, in contrast, means that followers agree with, accept and comply with the leader in exchange for praise, rewards and resources or the avoidance of disciplinary action. Building on these two notions Bass (1985) argues that transformational and transactional leadership are separate concepts and suggests that "the best leaders are both transformational and transactional" (Bass 1999, p. 21). Aviolo (1999) even suggests that "transactions are the base for transformations" (p. 37). Recent meta-studies on leadership styles seem to support this complimentary view that without the foundation of transactional leadership, transformational effects may not be possible (see for example, Lowe et al. 1996, Judge and Piccolo 2004, Bono and Judge 2004). In this study we will have a closer look at the two main tools available for transactional leadership: rewards and punishment. We will focus on the leader's choice problem as well on the effect on followers. In a literature review Podsakoff (1982) concludes that "research on the variables affecting a supervisor's use of rewards and punishment is still in its infancy" (p. 76). With few recent studies on this topic as notable exceptions his statement still appears to be valid. An important issue which needs further investigation in this respect is the conventional wisdom that leaders are very powerful in influencing the culture in a team or organization. The most prominent proponent of this view is Schein (2004) who argues that "I believe that cultures begin with leaders who impose their own values and assumptions on a group" (p. 2).[35] So far, the impact of the cultural variations induced by punishment and rewards combined with the respective incentive effects on success is still an open question. To shed light on this issue we investigate how a team leader chooses between positive and negative sanctions and how his choice influences the performance of teammates. Since these two questions can only be answered on an empirical basis we opt for an experimental approach, which has the decisive advantage that one can control for all situational variables and unambiguously observe the chosen actions.

In our simple model, the leader is a *primus inter pares*, i.e. he contributes to the team's production as the other team members do but he is the one who decides on the incentive

[33] This chapter is based on the working paper "Motivating Teammates: The Leader's Choice of Positive and Negative Incentives", joint work with Bernd Irlenbusch and Bettina Rockenbach. All authors contributed equally.

[34] Personnel Economics for Managers, 1998, pp. 269-270

[35] Schein (2004) defines culture of a group as "a pattern of shared basic assumptions that was learned by a group as it solved its problems of external adaptation and internal integration, that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems" (p. 17).

culture and the actual application of the specified sanctions.[36] The production process of the team is modeled such that from the perspective of an individual member it is beneficial to free-ride although from the viewpoint of the team as a whole everybody should contribute as much as he can. To overcome this dilemma situation the leader can apply sanctions. We assume that the leader is able to perfectly monitor the members of the team and has the power to apply sanctions to the other team members. Which sanctioning scheme do team leaders prefer in such a situation? How does the leader's choice influence the performance of the teammates? Do the prospects of receiving rewards motivate the teammates more than the threat of punishment? Which scheme induces more cooperation? Which is more profitable for the team, the leader, and the teammates?

In recent studies it has been observed that punishment fosters cooperation when it is available in a situation that is prone to free-riding (see e.g. Yamagishi 1986, Fehr and Gächter, 2000). Punishment may enhance cooperation even it is only symbolic (Masclet et al. 2003). However, punishment seems to have ambivalent effects on cooperation when it can deliberately be adopted by a party or peers before they interact with each other. While people seem to be reluctant to opt for negative sanctions (Sutter et al. 2006) and detrimental effects on cooperation in bilateral contracting have been observed (Fehr and Rockenbach 2003) there is also evidence that over time – after experiencing free-riding – punishment institutions becomes widely accepted and indeed improves cooperation level (Gürerk et al. 2007), negative sanctions can have. Dickinson and Isaac (1998) report that (exogenous) rewarding of both absolute and relative high contributions increase efforts towards a joint project. Contributions are highest when rewards are given for relative contributions, i.e., the contributions are evaluated with respect to the heterogeneous endowments ("abilities") of each member. Dickinson (2001) shows that exogenous penalties (negative incentives) are more efficient than prizes (positive incentives) when the contributions of heterogeneously endowed team members are evaluated after their absolute contributions to the public good. Sefton et al. (forthcoming) report a similar result from a public goods game. They find that the contributions and the earnings are higher when both reward *and* punishment possibilities are present. Sutter (2006) reports that in a threshold public good game a tournament design induces higher efforts than equally high rewards. Potters et al. (2005a) find that the presence of a leader who has discretionary power to determine the shares from the team output improves the team performance compared to a situation in which the team output is split among the team members equally.

The study by Sutter et al. (2006) is similar to ours with respect to the contribution mechanism and the endogenous choice of sanctions. In one of their treatment all players *vote* on a mechanism before interacting in a public goods setting (the alternatives include peer rewarding, peer punishment or simply no sanctioning). In our setting, only one single member – the leader of a team – chooses the incentive scheme. Second, in our setting, the leader is the only member of the team who has the possibility to reward or to punish while he himself is not subject to sanctions from none of the teammates. In our setting after a certain period of time the leader can change the institution, i.e., the available incentive scheme. Our study is also related to recent studies on leadership, in which the leader can give an example by choosing his contribution first which is then communicated to the followers before they decide on their contributions. It has been shown that in a setting with asymmetric information on the productivity of the team, followers mimic the informed leader's actions when contributing to a public good (Potters et al., 2005b, Potters, 2007). But also in situations with

[36] Alchian and Demsetz (1972) also suggest a team leader, who monitors the teammates, as a solution to the free-rider problem in teams. Their team leader, however, is the residual claimant, i.e., he is paid the residual of the team's profit minus the compensation of the teammates.

symmetric information second movers often follow the leader. Moxnes and van der Heijden (2003) find that the contributions to a "public bad" are lower when there is a leader who decides first compared to the situation without a leader and simultaneous decisions. A similar finding is reported by Gächter and Renner (2004). In a public good setting followers invest the more the more the leader contributes. This correlation is also true in a one-shot experiment. However, in the repeated interactions the followers contribute systemically less than the leader. Thus, leaders also reduce their contributions over time, which brings down the team production over time. Güth et al. (2007) find that leading by example induces a marginal but significant increase in contributions in a public good setting. They also show that if a leader has the power to (temporarily) exclude a team member contributions increase.

The paper is structured as follows. The next section introduces our experimental model and design. In section 3 we provide some thoughts on expected findings and section 4 reports the results. Section 5 concludes with some remarks on implications and suggestions for future research.

4.2. Experimental Design

We model team production in a voluntary contribution setting. A team consists of $N = 6$ members. The role of the *team leader* is randomly assigned to one of the team members. The team leader's role is identical to that of all other team members apart from his ability to choose an incentive scheme and his power to exert sanctions. The team leader chooses between a positive (POS) and a negative (NEG) incentive scheme, which will be applied to the next $T = 10$ periods.[37] All other N-1 team members, whom in the following we will refer to as *teammates*, are informed about the leader's decision regarding the incentive scheme. Each period consists of two stages. In stage 1 the team leader and the teammates decide simultaneously about the effort they want to contribute to their team's project, i.e., we basically model the team project by a voluntary contribution mechanism with a constant marginal productivity $R = 1.6$. For simplicity the costs of effort are assumed to be the same for all agents and equal to 1 for each effort unit. Effort per agent is restricted to a maximum of $y = 20$, i.e., $c_i(e_i) = e_i$ with $0 \le e_i \le y$ for $i = 1, \ldots, N$. If $q > 1$ is the exogenously given prize for one unit of output the total profit of the team is given by $qR(e_1 + e_2 + \ldots + e_N)$. Let $0 < \phi < 1$ denote the share of the team's profit that the firm gives to the team as wage. Note that we abstract from modeling the firm explicitly. If one assumes that the team members apply an equal sharing rule, each team member earns $\phi qR(e_1 + e_2 + \ldots + e_N)/N$. To keep things simple in the experiment we normalize ϕq to be equal to 1. Thus, the marginal per capita return (MPCR) a in our setting amounts to $a = (R/N) = (4/15)$ and satisfies the condition $1/N < a < 1$, which means that it is individually rational not to contribute to the team's output although it would be socially optimal to contribute maximal effort.[38] In stage 2, dependent on the chosen incentive scheme, the leader has the possibility to individually assign positive or negative tokens to the teammates. For this purpose the leader exogenously receives 20 additional tokens – one might think of an extra budget that is given from a higher management level to the team leader for the bonus payments. The leader can assign the additional tokens to the teammates or simply keep them as fringe benefits for the own account. Both positive and negative incentives have a leverage of 1:3, i.e., for each token invested by the leader, the payoff of the recipient is increased by 3 tokens in POS, or decreased by 3 tokens in NEG, respectively. At the end of each period, all team members

[37] We keep the incentives schemes fixed for 10 periods to resemble the fact that corporate codes of conducts and corporate cultures cannot be changed every day but have to be stuck to for a certain period of time (see Schein 2004).

[38] See Holmström (1982), for similar approaches to model team production in experiments see Nalbantian and Schotter (1997), Croson (2001), Sutter (2006) and Irlenbusch and Ruchala, forthcoming.

receive feedback about all individual contributions, payoffs and received tokens. The values for the leader and the teammates are indicated separately.

Our experiment consisted of three phases with $T = 10$ periods each, i.e., the leader decided at three points in time on the incentive scheme that ruled the next 10 consecutive periods. The team composition was fixed for the duration of the whole experiment. In Figure 4.1 the sequence of the experiment is visualized.

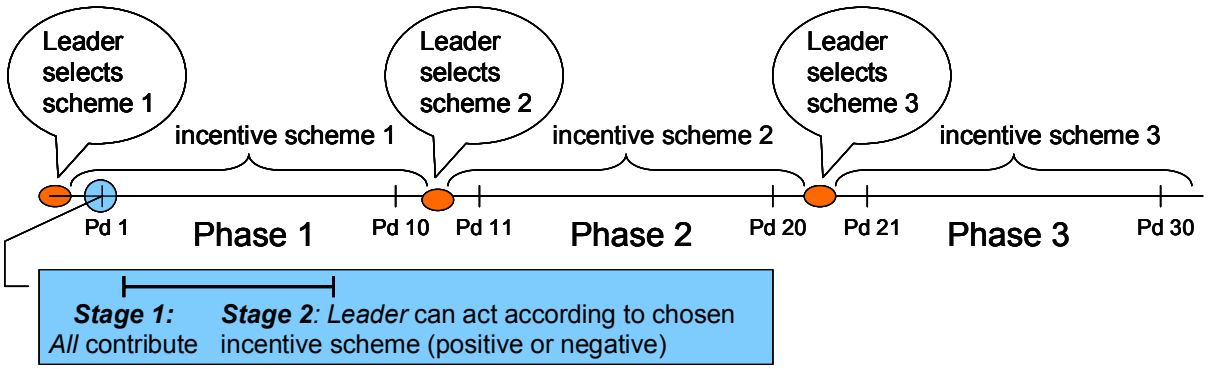


Figure 4.1 The sequence of actions in the experiment

60 students from the University of Erfurt were randomly allocated to 10 experimental teams of six subjects each, i.e., we have 10 independent observations. In the instructions we employed a neutral language. We labeled team leaders as type *A* players and the teammates as type *B* players. We did not speak of rewards or punishment. Instead, we used the terms positive and negative tokens that can be assigned by the type *A* player to type *B* players. A session lasted for approximately 90 minutes and average earnings were about 15 Euros.

4.3. Which incentive scheme will a leader choose?

A leader with *self-centered preferences* is only interested in maximizing his or her monetary payoff and should not engage in costly reward or punishment activities. Thus, in our setting the leader should be indifferent between both incentive schemes since with both incentive schemes he can keep the same amount of endowment.

A leader with *social preferences* does not solely care for the own monetary payoff. For example, an inequity averse player as defined by Fehr and Schmidt (1999) cares also about the inequality between the own and other players' payoff and both payoff inequalities – to the own advantage or to the own disadvantage – cause utility losses. Several researchers have already come up with theoretical explanations why teams, in which members exhibit social preferences, tend to behave more cooperatively (see e.g. Huck and Biel, forthcoming, Biel 2004, Mohnen et al. 2007). Their explanations are related to the effect of peer pressure which has also been verified empirically (see e.g. Falk and Ichino 2003). If negative sanctions are available Fehr and Schmidt have shown that also a single player with preferences of inequity aversion is able to discipline a whole group of free-riders by a credible threat to punish. Thus, in NEG a leader with a sufficient distaste for disadvantageous inequality in payoffs is potentially able to "enforce" a positive contribution level. In general this is less likely with positive sanctions as in POS, in which a leader with a distaste for disadvantageous inequality is not unilaterally able to force the teammates to contribute (see Gürerk et al. 2007). Thus, a leader who dislikes disadvantageous inequality is more likely to opt for NEG as an incentive scheme.

A leader with *efficiency preferences* who is interested in maximizing the total utility of the team might be inclined to choose POS since in POS the leader has the possibility to unilaterally increase the efficiency by allocating rewards. The reason is that each reward token assigned by the leader to a teammate costs the leader one token but increases the teammate's payoff by three tokens. Thus, by rewarding, the leader can increase the joint payoff of the team by two tokens. Several researchers have shown that efficiency is indeed an important driving force (see e.g. Charness and Rabin 2002, Huck et al. 2007 provide an illuminating approach by arguing that social norms inside the firm tend to develop into the direction to support efficiency).

4.4. Results

In this section we report our experimental findings. First we investigate the leaders' incentive scheme choices. We continue by comparing the leaders' and teammates' contributions dependent on the incentives schemes. Finally, we analyze the impact of positive and negative incentives which leads us to efficiency considerations. All reported non-parametric statistical tests are based on the averages over independent observations.

4.4.1. Leaders' incentive scheme choice

Table 4.1 gives an overview of the leaders' incentive scheme choices as well as the contributions and payoffs dependent on the player type. Overall, we observe five different realized sequences of choices. According to the choices in the first and the last phase of the experiment we can summarize these five different sequences in three patterns. Three leaders choose POS in the first and in the last phase. Six leaders choose also POS in the first phase while they choose NEG in the last phase. The remaining leader initially chooses NEG but opts for POS in the last two phases.

Table 4.1: Leaders' incentive scheme choices and average contributions

	Incentive scheme choice			Contributions						Payoffs						Efficiency in percent		
	Phase			Phase			Phase			Phase			Phase					
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Leader's ID				Leader			Teammates			Leader			Teammates					
5, 8	POS	POS	POS	11.2	11.0	8.9	13.2	12.1	8.6	35.5	36.2	37.5	35.7	34.0	29.7	38.7	37.4	33.7
1, 3, 6	POS	POS	NEG	7.6	7.6	9.3	8.8	10.5	14.5	35.3	35.2	42.2	31.5	33.5	21.1	34.9	36.7	47.3
2, 4, 9	POS	NEG	NEG	11.9	14.4	14	8.9	14.5	14.8	36.0	45.6	45.1	30.4	26.7	26.0	34.0	57.4	56.1
10	POS	NEG	POS	9.2	15.5	7.5	6.2	6.6	7.9	32.5	36.4	36.0	29.9	25.7	30.1	33.0	52.9	33.8
7	NEG	POS	POS	5.8	2.6	1.8	4.3	8.1	3.7	39.7	44.7	40.8	21.9	25.9	23.4	47.8	31.6	28.6

Efficiencies are given as percentages from the maximal obtainable total payoff.

In the initial phase, the overwhelming majority of team leaders (9 out of 10) opt for POS whereas only one leader prefers NEG. A binomial test rejects that the observed distribution of choices could come about by chance (p = 0.021, two-tailed). Hence, initially, leaders are clearly reluctant to adapt negative incentives for their teams.

Result 1 Initially, the large majority of leaders (90%) prefers the positive incentive scheme.

In phase 2, five leaders stick to POS whereas four leaders switch from POS to NEG. Interestingly, the only team leader who initially chooses NEG and switches to POS after the first phase. Thus, in phase 2, the majority of the leaders (60%) again opt for rewards. Why do some leaders change to NEG whereas others prefer to stick to POS? Three of the four leaders who change from POS to NEG contribute much more to the public good than their teammates

before the change (on average 64.9% more). On the other hand, three of the five leaders, who stick to POS in phase 2, contribute considerably less in phase 1 than their teammates (about 27% less). Thus, leaders who switch in phase 2 from POS to NEG are likely to be motivated by the disappointingly low contributions of teammates compared to their own contribution.

In the last phase of the experiment, three of the four leaders who choose NEG in phase 2 remain in NEG. Only one leader (observation 10) switches from NEG in phase 2 to POS. Thus, leader 10 is the only one showing a back and forth behavior choosing initially POS, switching to NEG in phase 2 and finally turning back to POS in phase 3. Three of the five leaders who choose POS twice switch to NEG in phase 3. Hence, the majority of leaders (60%) opt for NEG in the last phase.

Result 2 In the last phase of the experiment, the majority of leaders (60%) opt for the negative incentive scheme.

4.4.2. Evolution of contributions

Figure 4.2 shows the development of average team contributions in different phases under both incentive schemes.[39] What is the reason for the higher contributions in NEG? Is it because teammates fear to be punished? Do average contributions by leaders – who themselves cannot be punished – differ from their teammates after a change from POS to NEG? Overall, 6 of 7 leaders who change from POS in phase t to NEG in phase t+1 also significantly increase their average contributions in phase t+1 (Wilcoxon test, $p = 0.047$). In the initial phase, teams under POS contribute more than the single group using NEG. However, already in phase 2, teams contribute higher in NEG than in POS.

Result 3 Overall (average of phases 1-3), team contributions are higher under NEG than under POS (Wilcoxon test, $p = 0.016$). This is also true if one considers the contributions of leaders and teammates separately (Wilcoxon test, $p = 0.040$, both).

Moreover, in *none* of the seven cases, where a team switches from POS to NEG, the *relation* changes between the contributions of the leader and the teammates. This means that, if a leader contributes on average more than his teammates before changing from POS to NEG (this is the case in 4 teams) the leader continues to contribute more after the change in NEG. Analogously, if a leader contributes on average less than the teammates before a change from POS to NEG (in the remaining 3 teams), the leader also contributes less after the change. Interestingly, in the two cases where leaders switch in the opposite direction, i.e., from NEG to POS, the relation between the leaders' and the teammates' contributions reverses after the change, i.e., in these cases, leaders contribute more than their teammates before the change, however, after the change teammates contribute more than their leaders.

Result 4 A change from positive to negative incentives does not alter the relation between the leader's and the teammates' contributions.

[39] For an overview of different contributions over periods in different teams see appendix Figure A4.1. Average path-dependant contributions in different phases are provided in Figure A4.3.

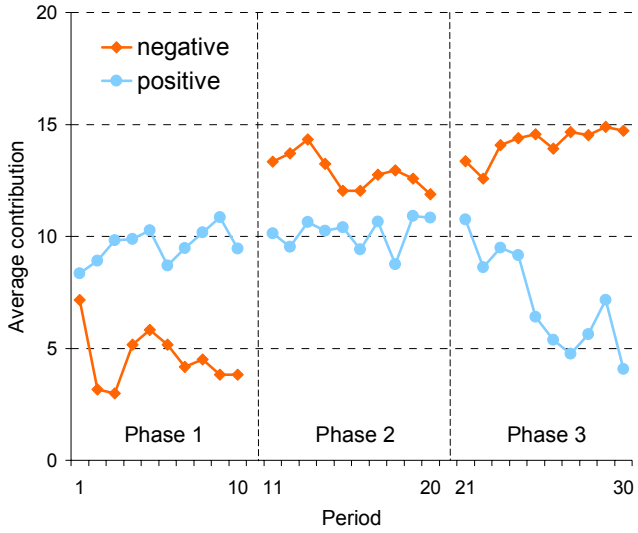


Figure 4.2 Team contributions dependent on the incentive scheme

Does switching from POS to NEG have an immediate effect on contributions? To answer this question we compare the last period contributions in POS (periods 10 and 20, respectively) before the change with the first period contributions in NEG after the change (periods 11 and 21, respectively). In total, seven teams change from POS to NEG. In *each* of these teams contributions rise in the first period after change from on average 8.8 to on average 12.8 (Wilcoxon test, $p = 0.016$). Interestingly, *both* leaders and teammates increase their contributions immediately. On average, leaders' contributions rise from 9.9 to 12.4 units whereas the increase in teammates' contributions is even larger from 8.6 to 12.8 units (the latter is significant, Wilcoxon test, $p = 0.016$; the former turns out not to be significant which is mainly because two leaders are not able to increase their contributions since they already contribute 20 before the change; in fact only one leader contributes less after a change to NEG).

The effect on contributions of choosing POS two times in a row is not so clear. In total we have seven teams in which a leader chooses POS again after having been in POS before. On average, teammates decrease their contributions immediately in the first period of the second POS stage (from 10.9 to 10.5 units). Leaders, however, increase their contributions slightly (from 6.5 to 6.8 units). Both differences are not statistically significant, however.

A change from NEG to POS occurs only in two teams. In both cases teammates increase their contributions while one leader increases and the other decreases their contributions. Due to the small number of observations, we cannot say much on the effect of this type of change.

The question arises whether the teammates follow the leader's example in the sense that they adjust their contributions towards the contribution of the leader. Figure 4.3 a) and b) show the evolution of the contributions dependent on player type and incentive scheme. Frequency distributions are shown in 3c) and d).

On average, there exists no significant difference between leaders' and teammates' contributions under both incentive schemes. This is somewhat surprising since the leader is the only team member who cannot be sanctioned and thus might be less motivated to contribute. Does the leader exert higher effort to give an example for the other team members? In our setting the contribution decisions are made by all team members simultaneously. Thus, it is not possible that the team members follow their leader already in

the same period. However, imitation of the leader by the teammates may take place with a time lag, i.e., teammates may adapt to the leader's contribution from the previous period (which they are informed about after each period). To investigate whether such a following behavior occurs in our experiment we run an interval regression. We ask whether the difference between the leader's and a teammate's contribution in period *t*-1 influences the contribution of this teammate in period *t*. As can be seen in Table 4.2, under both incentive schemes, the leader's contribution in the current period indeed has a positive impact on teammates' contributions in the following period. The actual influence of the leader's contribution in the previous round is smaller in NEG than in POS. However, the effect of the sanctioning mechanism is considerable higher in NEG than in POS. Actually, the effect of positive sanctions turns out to be negative. This is in line with findings from Gürerk et al. (2006). We will discuss possible interpretations below.

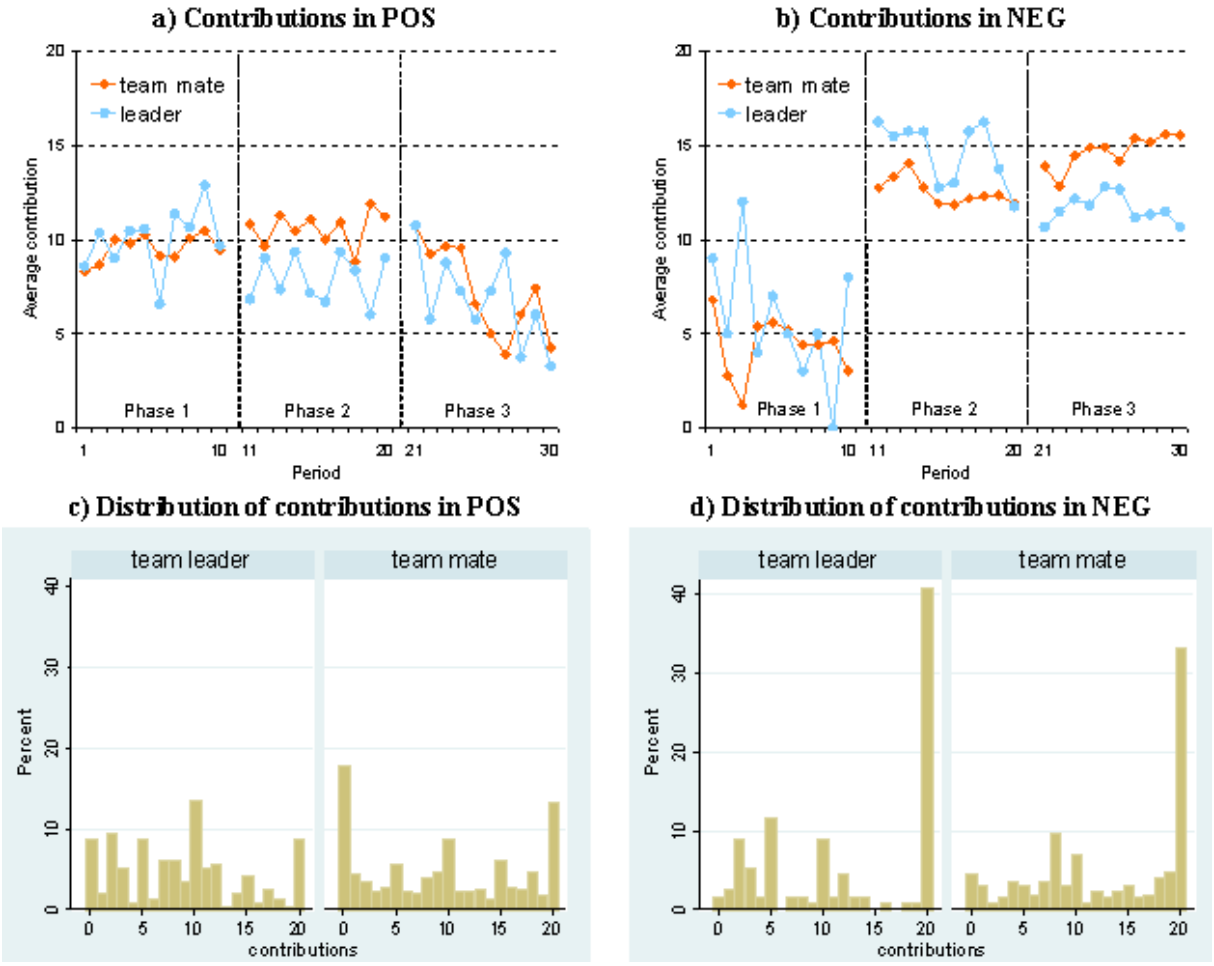


Figure 4.3 Team contributions dependent on type & incentive scheme

Table 4.2: Teammates' contributions explained by differences between own and leader's contributions

Coefficient	POS	NEG
Leader's contribution (in *t*–1) – teammate's contribution (in *t*–1)	0.34***	0.08*
	(0.07)	(0.04)
Received tokens in (*t*–1)	-0.29*	0.84***
	(0.16)	(0.22)
Constant	0.72**	-0.84***
	(0.31)	(0.26)

***, **, and * denote significance level 0.01, 0.05, and 0.1 respectively.
Robust standard errors in parentheses (adjusted for 19 clusters in POS and for 11 clusters in NEG).

4.4.3. Use of incentives

When do team leaders actually use the available incentive mechanism? Figure 4.4a) shows the percentage of cases where a teammate receives positive or negative tokens from the leader depending on the difference between own teammate's and leader's contributions. The overwhelming majority of 88% of those teammates who outperform leaders' contributions are rewarded. On average, they receive 3.4 tokens, i.e., their payoff is increased by 10.2 tokens. Leaders also appreciate teammates who contribute exactly the same amount as they themselves (66% of these teammates are rewarded). They receive 3.1 tokens on average (i.e. their payoff is increased by 9.3 tokens). Surprisingly, roughly 40% of teammates who contribute less than their team leaders are also rewarded which might be interpreted as an encouragement for exerting higher efforts in the future or just as a means to increase efficiency. On average they receive 1.8 tokens on average (see Figure 4.4b), i.e., their period payoff is increased by 5.4 tokens on average. Teammates who free-ride completely, i.e., who contribute less than 5, are not rewarded at all. Thus, the use of positive incentives increases monotonically with the contribution difference between a teammate and the leader.

Result 5 The frequency and the extent of positive incentives increases with the contribution difference between the leader and the respective teammate. Teammates who contribute the same as the leader are also considerably rewarded.

In nine of the ten independent observations the coefficients of a Spearman-rank-correlation between contribution differences and reward tokens are highly positive. Hence a binomial test with these correlation coefficients clearly rejects that a positive correlation is as likely as a negative correlation ($p = 0.021$).

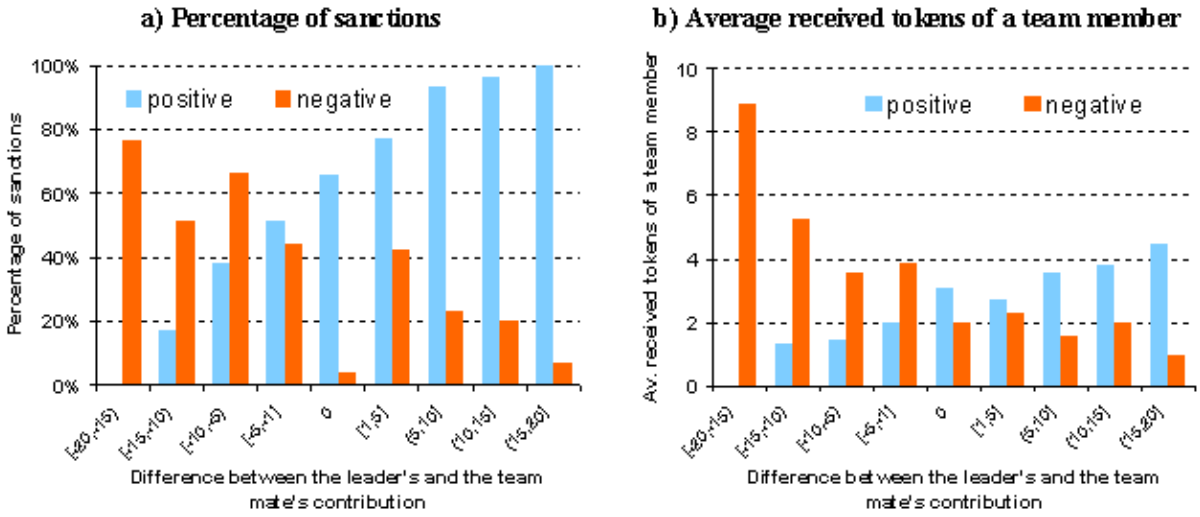


Figure 4.4 Frequency and extent of sanctions

Now we turn to the use of negative incentives. Leaders punish free-riders frequently and harshly. Almost 80% of free-riders are punished. On average, free-riders receive 8.9 negative tokens, i.e., their period payoff is reduced by 26.7 tokens. In total, 54% of teammates who contribute less than the leader receive negative tokens. Teammates who contribute exactly the same as the leader are virtually never sanctioned (4%). Surprisingly, leaders assign negative tokens also to teammates who contribute more than they do themselves. In fact this is true in 28% of the cases in which teammates contribute more than the leader.

Result 6 The frequency and the extent of negative incentives decreases with the contribution difference between the leader and the other team members. Teammates who contribute exactly the same amount as the leader are virtually never sanctioned.

In each of the eight independent observations where negative incentives are chosen the coefficients of Spearman-rank-correlations between contribution differences and punishment tokens are highly negative. A binomial test with these correlation coefficients clearly rejects that a negative correlation is as likely as a positive correlation ($p = 0.008$). Table 4.3 provides Tobit regressions which explain the allocated sanctioning tokens by the contribution of the respective teammate and the difference between his contribution and that of the leader.

Table 4.3: Leaders' choice of sanctioning tokens explained by teammates' contributions and contribution difference to the leader

Coefficient	POS	NEG
Teammate's contribution in *t*	0.28***	-0.31***
	(0.01)	(0.05)
Leader's contribution in *t* – teammate's contribution in *t*	-0.04***	0.17***
	(0.01)	(0.04)
Constant	-1.26***	1.29**
	(0.13)	(0.61)

***, and ** denote significance level 0.01, and 0.05 respectively.

What is the effect of incentives on the contributions of teammates? The regression provided in Table 4.2 already indicates that negative sanctions have an unambiguously positive influence on contributions while the influence of positive sanctions seems in fact to be negative. This is confirmed by Figure 4.5 which shows the immediate average change in teammates' contributions from period *t* to period *t*+1 after receiving positive or negative tokens in period *t*. On average, teammates react to negative incentives with an increase of contributions. The more negative tokens a teammate receives the higher is the increase in contributions. On the other hand, teammates who do not receive any negative tokens in period *t*, reduce on average their contributions in period *t*+1 roughly by one token.

Positive incentives have a different effect on contributions. On average, teammates decrease their contributions when they receive rewards. Roughly speaking it is true that the more positive tokens are received the stronger is the decrease in contributions. Not rewarding a teammate has also an effect. Teammates who do not receive any positive tokens increase their contributions in the following period almost by four tokens on average.

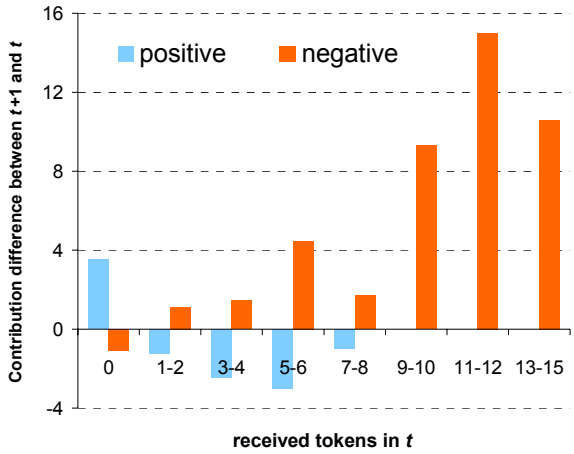


Figure 4.5 The effect of sanctions on the contribution change of teammates

4.4.4. Payoffs and Efficiency

Overall, team payoffs are higher in POS than in NEG (Wilcoxon test, $p = 0.008$). Note that the positive tokens sent by the leaders are tripled and thereby positive sanctions increase overall efficiency. Nevertheless, the differences in payoffs between both schemes diminish in the last phase. In the very last period due to high contributions the payoffs in NEG are even higher than the payoffs in POS (see Figure 4.6).

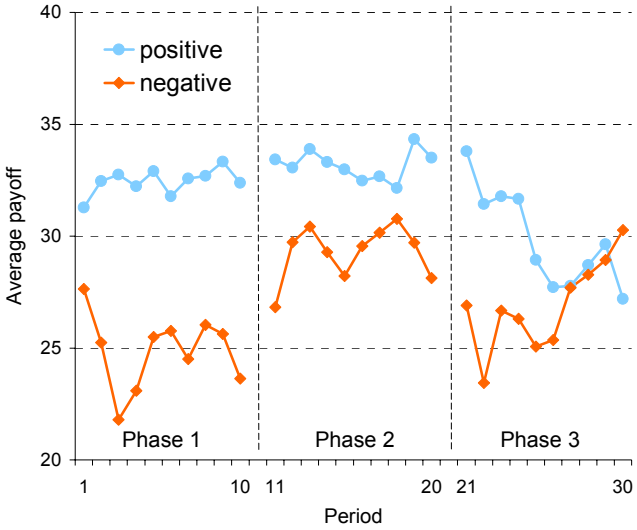


Figure 4.6 Team payoffs dependent on the incentive scheme

In both schemes, leaders' payoffs are significantly higher than teammates' payoffs (see Figure 4.7). This is true for POS (Wilcoxon test, $p = 0.028$) as well as for NEG (Wilcoxon test, $p = 0.008$). However, the differences between the payoffs are less pronounced in POS. When comparing the payoffs of both types of players it turns out that teammates' earnings are higher in POS than in NEG (Wilcoxon test, $p = 0.040$) whereas leaders earn more in NEG than in POS (Wilcoxon test, $p = 0.110$).

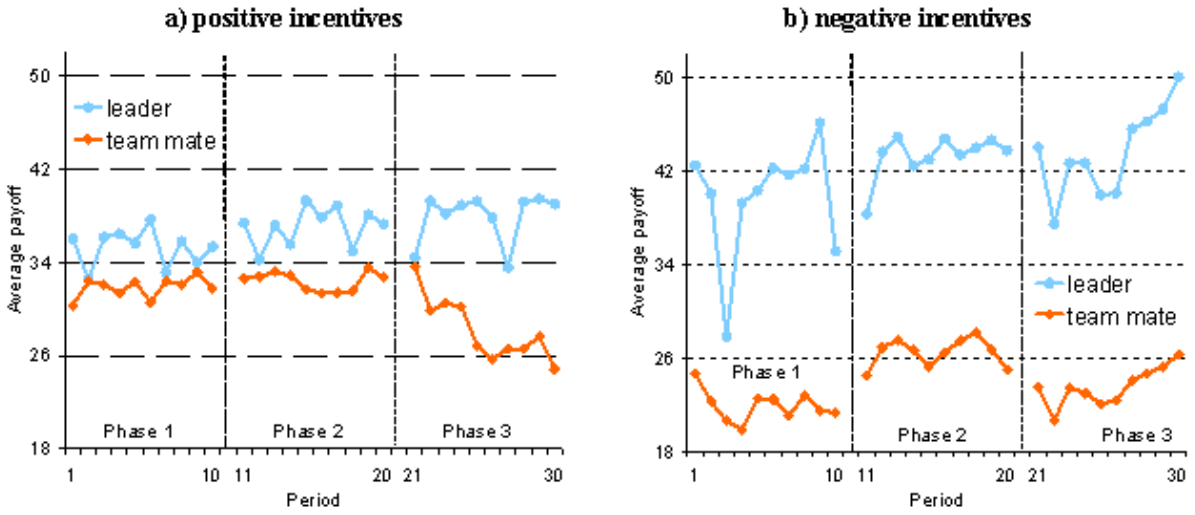


Figure 4.7 Team payoffs dependent on type & incentive scheme

Leaders in NEG obtain higher payoffs for two reasons. As can be seen in Figure 4.8, leader payoffs from both stages are higher in NEG than in POS. This means that the leaders' returns from the team production are higher in NEG than in POS in essence because average contributions – by leaders *and* by teammates – are higher in NEG. Moreover, in stage 2

leaders spent less on negative incentives in NEG than they spent on the positive incentives in POS. Also teammates' payoffs from the team production stage are higher in NEG than in POS. However, in stage 2 teammates in NEG experience a decrease in payoffs due to received negative tokens while their payoff increases in POS due to the received positive tokens. The net effect from both stages is that teammates earn less in NEG than in POS although contributions are higher in NEG.

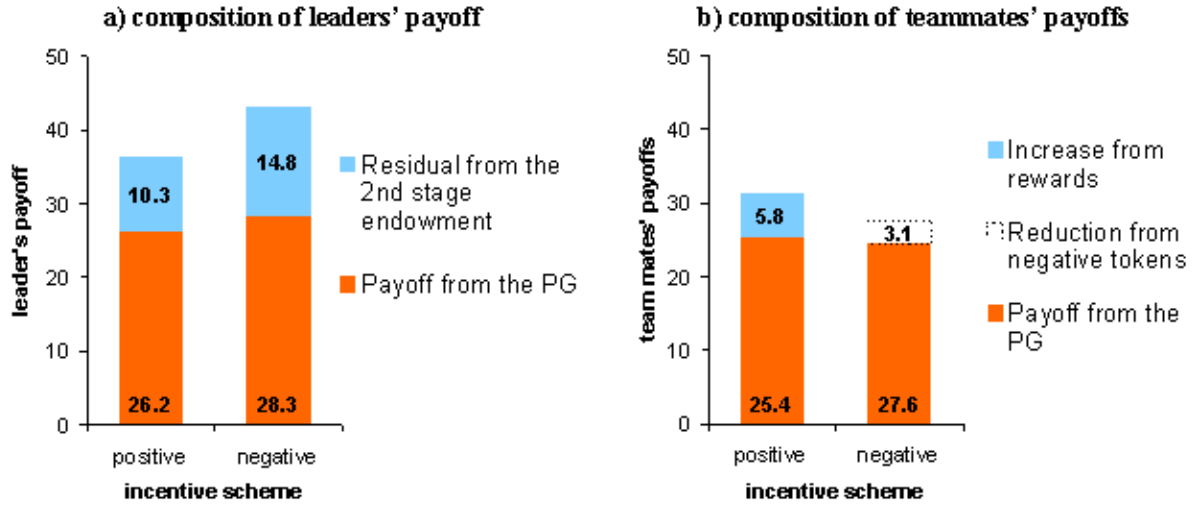


Figure 4.8 Composition of payoffs

We measure efficiency as the relation between the possible maximum payoff a team can obtain and the payoff a team actually achieves in the experiment. As already argued above, in POS, it is socially optimal (= maximum joint payoff) if the leader allocates all reward tokens to the teammates since every token sent by the leader increases the team's net payoff by two additional tokens. In NEG, however, it is socially optimal when the team leader refrains from assigning any tokens to the teammates since each token in total reduces the team payoff by four tokens. In NEG, teams obtain 76.0% of the maximum possible payoff whereas in POS teams only manage to obtain 75.8% of the maximum possible payoff. Thus, both incentive schemes do not differ significantly from each other in terms of efficiency achieved.

4.5. Conclusion

In this study we report on an experiment designed to analyze the behavior of a team leader who can choose a leadership style which relies either more on punishment or more on rewards as an incentive mechanism. Both mechanisms are seen as essential ingredients for successful transactional leadership. The question is, however, how a predominant focus on one of these mechanism influences a team's performance. The choice of the leader is known to the team members before they decide on their contributions and it is kept fixed for a certain period of time. We keep the incentive scheme fixed for some periods (before it can be altered again) to account for the facts that an established culture in a team or organization – which is largely shaped by the prevailing incentive mechanism – cannot be changed on a day by day basis. The induced culture might indeed influence the contribution behavior of organizational members. A constant threat of punishment might be perceived as discouraging by the subordinates while a constant need for reward might burden the leader.

We find that the overwhelming majority of 90 percent of the leaders opt for the positive incentives in phase 1. This finding is in line with the observed reluctance regarding the punishment option observed in other studies (Botelho et al. 2007, Sutter et al. 2006). However, the initial preference for rewards changes during the experiment since in the last phase, 60 percent of the leaders choose the negative incentives. This reflects the findings

reported in Gürerk et al. (2006). Overall contributions are higher if the negative sanctions are employed than in presence of rewards. Interestingly, this is not only true for the teammates but also for the leader who cannot be punished. A change from positive incentives to negative ones results in an immediate increase in contributions, i.e., the anticipation of potential punishment has already a positive effect on contributions. However, if the leader contributed already more before the switch he also does so afterwards. The application of negative and positive sanctions can largely be explained by the absolute amount of the contribution of the respective teammate and the contribution difference between him and the leader. While negative sanctions have an unambiguously increasing effect on contributions this is not the case for rewards. We can only speculate why this is the case. One explanation could be that teammates perceive a reward as a signal that they actually contributed more than was expected by the leader and therefore slack off afterwards. It appears to be quite promising to investigate this puzzling observation in more detail in the future. Another observation which deserves further investigation is that some leaders punish even if the respective team member contributed more than the leader himself. This finding is in line with observations by Cinyabuguma et al. (2006) who report frequent punishment of high contributors in standard VCMs. Apparently some leaders are not satisfied with or simply do not agree with contributions that exceed their own contributions. Neither standard economic theory nor inequality aversion can explain such behavior since assigning negative tokens to high contributors is costly and does not reduce the inequality in payoffs. On the contrary, it enlarges the inequality to the disadvantage of the teammates. We also observe a "leading by example effect" reported already in previous studies, i.e., teammates seem to contribute more the more the leader contributes (compare Gächter and Renner 2004, Güth et al., 2007). Surprisingly, this effect seems to be larger in a culture with positive sanctions than with negative ones. Apparently teammates are more reluctant to follow the leader if he threatens to punish. To investigate this effect in more detail also needs further research.

Of course, one has to be cautious with immediately transferring the findings from a lab experiment to the real situation of a team in an organization. However, we think that – in addition to the behavioral tendencies outlined above – the discovered trade-off between punishment and rewards is likely to be also relevant in real teams. On the one hand negative sanctions seem to be the more powerful tool to encourage team members to exert effort. On the other hand a leader might want to take advantage of the efficiency increasing effect of positive sanctions.

4.6. References

Alchian, A. and Demsetz, H. (1972). "Production, Information Costs, and Economic Organization." *American Economic Review*, 62, pp. 777-795.

Aviolo, B.J. (1999). "Full leadership development. Thousand Oaks, CA: Sage.

Bass, B.M. (1985). "Leadership and performance beyond expectations." New York: Free Press.

Bass, B.M. (1999). "Two decades of research and development in transformational leadership. *European Journal of Work and Organizational Psychology*, 8, 9-32.

Biel, P.R. (2004). "Inequity Aversion and Team Incentives." Discussion Paper, University College London.

Bono, J.E. and Judge, T. A. (2004). "Personality and Transformational and Transactional Leadership: A Meta-Analysis." *Journal of Applied Psychology*, 89(5), pp. 901-910.

Botelho, A., Harrison, G., Costa Pinto, L.M., and Rutström, E.E. (2007). "Social Norms and Social Choice." Working Paper.

Burns, J.M. (1978). "Leadership." New York: Harper & Row.

Charness, G., Rabin, M. (2002). "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117, pp. 817-869.

Cinyabuguma, M., Page, T., and Putterman, L. (2006). "On Perverse and Second-Order Punishment in Public Goods Experiments with Decentralized Sanctioning." *Experimental Economics*., 9, pp. 265–279.

Croson, R. (2001). "Feedback in Voluntary Contribution Mechanisms: An Experiment in Team Production." *Research in Experimental Economics*, 8, pp. 85-97.

Dickinson, D.L. and Isaac, R.M. (1998). "Absolute and Relative Rewards for Individuals in Team Production." *Managerial and Decision Economics*, 19, pp. 299–310.

Dickinson, D. L. (2001). "The Carrot vs. the Stick in Work Team Motivation" *Experimental Economics*, 4(1), pp. 107-124.

Falk, A. and Ichino, A. (2003). "Clean Evidence on Peer Effects." *Journal of Labor Economics*, 24(1), pp. 39-57.

Fehr, E. and Gächter, S. (2000). "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90(4), pp. 980-994.

Fehr, E. and Schmidt, K. (1999). "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, 114 (3), pp. 817-868.

Fehr, E. and Rockenbach, B. (2003). "Detrimental effects of sanctions on human altruism." *Nature*, 422 (6928), pp. 137-140.

Gächter, S. and Renner, E. (2004). "Leading by Example in the Presence of Free Rider Incentives." Working Paper, University of Nottingham.

Gürerk, Ö., Irlenbusch, B., and Rockenbach, B. (2006). "The Competitive Advantage of Sanctioning Institutions." *Science*, 312, pp. 108-111.

Gürerk, Ö., Irlenbusch, B., and Rockenbach, B. (2007). "Community Choice in Social Dilemmas – A Voting With Feet Approach." Working Paper, University of Erfurt.

Güth, W., Levati, M.V., Sutter, M. and van der Heijden, E. (2007). "Leading by example with and without exclusion power in voluntary contribution experiments." *Journal of Public Economics*, 91, pp.1023-1042.

Holmström, B. (1982). "Moral Hazards in Teams." *Bell Journal of Economics*, 13, pp. 324-340.

Huck, S. and Biel, P.R. "Endogenous leadership in teams." forthcoming in *Journal of Institutional and Theoretical Economics*.

Huck, S., Kübler, D. and Weibull, J. (2007). "Social Norms and Economic Incentives in Firms." Discussion Paper, University College London.

Irlenbusch, B. and Ruchala, G. "Relative Rewards Within Team-Based Compensation." forthcoming in *Labour Economics*.

Judge, T. A. and Piccolo, R.F. (2004). "Transformational and Transactional Leadership: A Meta-Analytic Test of Their Relative Validity." *Journal of Applied Psychology*, 89(5), pp. 755-768.

Lazear, E. (1998). "Personnel Economics for Managers." New York: Wiley.

Lowe, K.B., Kroeck, K.G., and Sivasubramaniam, N. (1996). "Effectiveness Correlates of Transformational and Transactional Leadership: A Meta-Analytic Review of the MLQ Literature." *Leadership Quarterly*, 7(3), pp. 385-425.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review*, 93(1), pp. 366-380.

Mohnen, A., Pokorny, K. and Sliwka, D. (2007). "Transparency, Inequity Aversion and the Dynamics of Peer Pressure in Teams – Theory and Evidence." Discussion Paper, University of Bonn.

Moxnes, E. and van der Heijden, E. (2003). "The Effect of Leadership in a Public Bad Experiment." *Journal of Conflict Resolution*, 47, pp. 773-795.

Nalbantian, H. R. and Schotter, A. (1997). "Productivity under Group Incentives: An Experimental Study." *American Economic Review*, 87, pp. 314-341.

Podsakoff, P.M. (1982). "Determinants of a Supervisor's Use of Rewards and Punishment. A Literature Review and Suggestions for Future Research." *Organizational Behavior and Human Performance*, 29, pp.58-83.

Potters, J., Sefton, M., and van der Heijden, E. (2005a). "Hierarchy and opportunism in teams." CentER Discussion Paper 2005-109.

Potters, J., Sefton, M., and Vesterlund, L. (2005b). "After you—endogenous sequencing in voluntary contribution games." *Journal of Public Economics*, 89, pp. 1399–1419.

Potters, J., Sefton, M., and Vesterlund, L. (2007). "Leading-by-example and signalling in voluntary contribution games: an experimental study." *Economic Theory*, 33, pp. 169–182.

Schein, E.H. (2004). "Organizational Culture and Leadership." John Wiley & Sons, Inc.: San Fransisco.

Sefton, M., Shupp, R., and Walker, J. "The Effect of Rewards and Sanctions in the Provision of Public Goods." Forthcoming in *Economic Inquiry*.

Sutter, M. (2006). "Endogenous versus exogenous allocation of prizes in teams—Theory and experimental evidence" *Labour Economics* 13, pp. 519–549.

Sutter, M., Haigner, S., and Kocher, M.G. (2006). "Choosing the carrot or the stick? – Endogenous institutional choice in social dilemma situations." CEPR Discussion Paper No. 5497.

Yamagishi, Toshio (1986). "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology* 51(1), pp. 110-116.

Appendix

Appendix 1.1. Materials and Methods

84 undergraduate students from the University of Erfurt voluntarily participated in the experiments. Special care was exerted to recruit students from many different disciplines to increase the likelihood that the subjects had never met before. Each participant was allowed to take part in one session only. In total 7 experimental sessions each involving 12 subjects took place. These sessions constituted the independent observations for the non-parametric statistical analysis. Most of the sessions were run in pairs, i.e., 24 subjects were gathered in the lab.

The game is repeated over 30 periods and participants are not restricted by choices performed in previous periods. Each period consists of three stages: An institution choice stage (S0), a voluntary contribution stage (S1), and a sanctioning stage (S2). In stage S0 the participants simultaneously and independently choose between a sanctioning institution (SI) and a sanction-free institution (SFI) in which neither positive sanctioning (rewards) nor negative sanctioning (punishment) is possible. In stage S1, each participant is informed about the number of participants in each institution and in case the institution is occupied by at least two participants a public goods game is played with all participants who have chosen the same institution in S0. If only one subject joins an institution the subject's total endowment is automatically transferred to her/his private account. The public good's game constitutes a prototypical social dilemma in which each player is endowed with 20 money units (MUs) and may contribute between 0 and 20 MUs to a project which benefits the entire group. Each MU contributed to the public good is deducted from the contributor's private account and creates a benefit of 1.6 MUs for the entire group. This group benefit is equally distributed among the group members, i.e. if a group consists of n members each member profits by $1.6/n$ MUs from each 1 MU contributed ($1.6/n$ is the marginal per capita return MPCR). If, for example, only one group member contributes the total endowment of 20 and the other $n-1$ group members contribute nothing, the public good amounts to $20 \cdot 1.6$ and the contributor's profit is $20 \cdot 1.6/n$ while each free-riders' profit is $20+20 \cdot 1.6/n$. If, however, all n group members contribute an identical amount of x, with $0 \leq x \leq 20$, the public good is $n \cdot x \cdot 1.6$ and each member achieves a profit of $20-x+1.6 \cdot x = 20+0.6 \cdot x$. Hence for an identical contribution x of all group members the net benefit of each group member is $0.6 \cdot x$ independent from the group size n. The MUs not contributed to the public good are transferred to the participant's private account. Thus, the provider's return from one additional MU is less than 1 but the group's return exceeds 1. Since the cost of providing is higher than the individual return, it is always in the material self-interest of any subject to free-ride on the contributions of the others and to keep all MUs for the private account. If all participants follow their material self-interest, nobody contributes to the public good and each participant achieves a payoff of 20 MUs. Because the group's return of each MU invested is greater than 1, it is in the collective interest that all group members contribute their entire endowment to the group project. These diametrically opposed individual and collective interests constitute the social dilemma in public good provision. After the players have simultaneously made their contribution decisions, they are informed about the contributions of each member in the own group.

At the beginning of stage S2 each player receives additional 20 tokens independent of the affiliation choice in S0. In SFI these tokens are directly transferred to the player's private account without any decisions required, i.e. sanctioning was not possible. In SI these tokens may be used to positively or negatively sanction other members of SI by assigning between zero and 20 tokens to other members. Each player is free to choose which of the other

members of SI she/he wants to positively and/or negatively sanction and to determine the amount of allocated sanctioning tokens to each of those players. She/he is free to allocate different numbers of sanctioning tokens to different individuals with the only restriction that the sum of allocated tokens is limited to at most 20. Tokens not used for sanctioning are transferred to the player's private account. Each token employed as a negative sanction costs the punished member 3 MUs and the punishing member 1 MU. Each token employed as a positive sanction yields the receiving member 1 MU and costs the employing member 1 MU. The leverage in the negative sanctioning mechanism is motivated by the understanding that punishment is more costly for the punished individual than for the punisher. We assume that the leverage in positive sanctioning is smaller and does not create any efficiency gains. The efficiency loss of negative sanctioning as well as the efficiency neutrality of positive sanctioning excludes efficiency gains solely by applying these instruments.

At the end of the period each participant receives detailed (but anonymous) information about each individual other participant from both institutions: the contribution, the sum of allocated positive sanctioning and negative sanctioning tokens to others, the sum of received positive sanctioning tokens from others, the sum of received negative sanctioning tokens from others, and the period profit. Players are neither informed about the identities of the other players nor are they able to track the identities over periods, because the order in which the players' details are displayed is known to be randomized in each period. In particular players could not identify the other players who allocated sanctioning tokens to them.

At the beginning of the experiment subjects received written instructions (see the section "Experimental instructions" below). At the end of the experiment subjects privately received their experimental earnings in cash. One experimental session typically lasted for 2.5 hours, and on average subjects earned 24 € per session. All experimental decisions were made on a computer screen using the experimental software z-Tree (Fischbacher, 1999). Each of the 24 computers was located in a booth such that subjects could not see or communicate with each other.

The Effect of the Group Size on the Marginal per Capita Return (MPCR)

The marginal per capita return (MPCR) denotes the individual return a recipient obtains from each token contributed. In a public goods game of n players the MPCR is lower than 1 and exceeds $1/n$. The MPCR being lower than 1 implies that it is individually rational to refrain from contributing since the individual return is lower than the investment. The fact that the MPCR exceeds $1/n$ implies that contributing is collectively rational because the groups' return on each token invested is greater than one. The endogenous group choice in each period of our experiment allows varying group sizes in each period. We constructed the MPCR such that it changes with the group size n, i.e. $\text{MPCR} = 1.6/n$. Hence, smaller groups have a higher MPCR than larger groups. Thus, the more members choose an institution the lower is the individual return on investment for one contributed token. As a consequence, however, the total "productivity" $n \cdot \text{MPCR}$ from the perspective of the group is constant, i.e. equal to 1.6. This means that all groups consisting of full cooperators achieve the same individual payoffs (i.e. $20 \cdot 1.6$), no matter how large the groups are. Hence, in the Nash-equilibrium of complete free-riding as well as under full cooperation the individual payoffs do not depend on the group size. From what is known on the interplay of the group size and the MPCR (Isaac et al, 1984) our design favors cooperation in small groups and disfavors cooperation in large groups.

Appendix 1.2. Supporting Figures and Tables

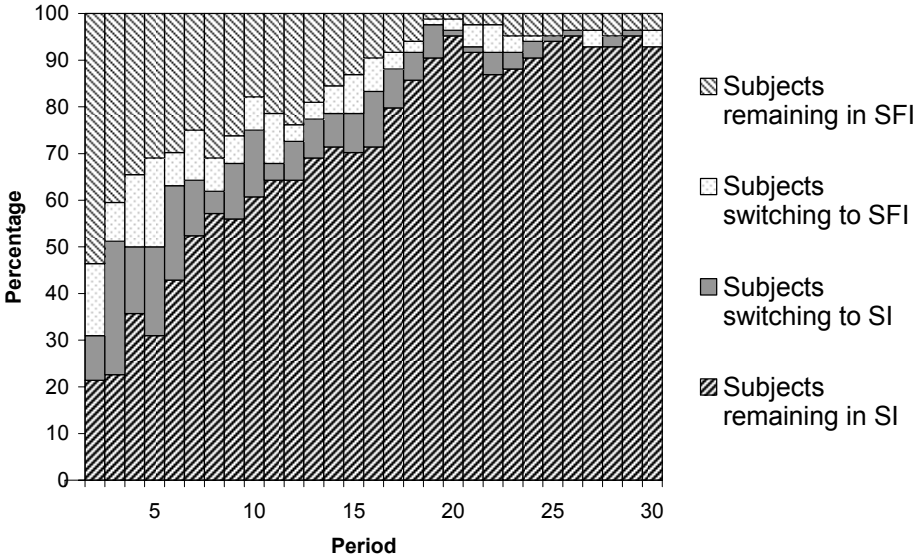


Figure A1.1 Subjects' choices of institutions and their switching behavior in both directions

Figure A1.1 displays the percentage of subjects who remain in the institutions or switch between the institutions.

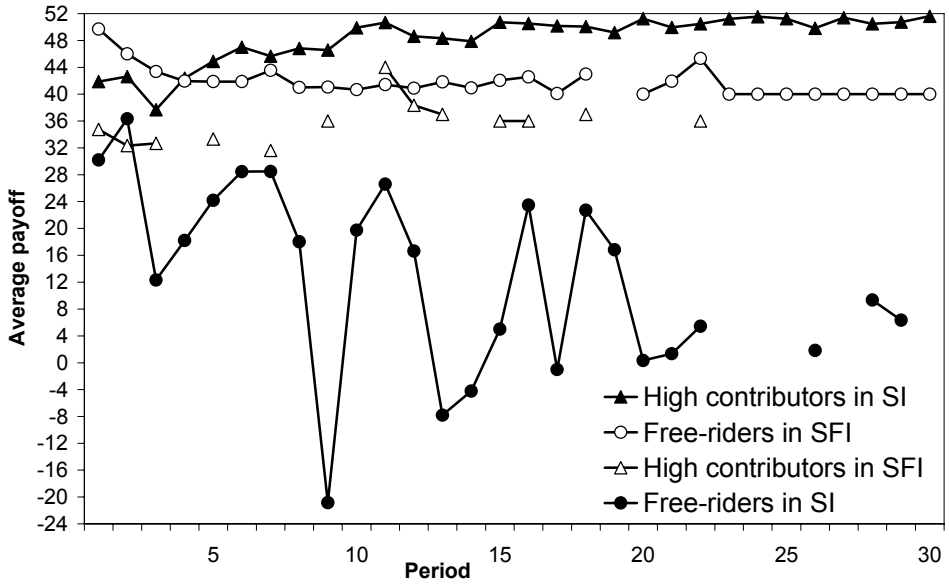


Figure A1.2 Subjects' average payoffs dependent on their contribution behavior

Table A1.1: Logit analysis of the punishment probability dependent on the experience in SI (using individual dummies for subjects, 778 observations in total)

	Coefficient		Z value
Number of periods in SI when punishing	-.074	(.026)	- 2.82***
Constant	.798	(.278)	2.87***

*** denotes significance at 1%. The values in parentheses denote the standard errors.

References to Appendix 1.2.

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

Isaac, M., Thomas, S., and Walker, J. "Divergent Expectations on Free Riding: An Experimental Examination of Possible Explanations." *Public Choice*, 1984, 43, pp.113-149.

Appendix 1.3. Instructions to the experiment

General Information: At the beginning of the experiment you will be randomly assigned to one of **2 subpopulations each consisting of 12 participants**. During the whole experiment you will interact only with the members of your subpopulation. At the beginning of the experiment, **1,000 experimental tokens** will be assigned to the experimental account of each participant.

Course of Action: The experiment consists of **30 rounds**. Each round consists of 2 stages. In Stage 1, the group choice and the decision regarding the contribution to the project take place. In Stage 2, participants may influence the earnings of the other group members.

Stage 1

(i) The Group Choice: In Stage 1, each participant decides which group she wants to join. There are two different groups that can be joined:

		Influence on the earnings of other group members
Group	A:	No
	B:	Yes, by assigning positive and negative tokens

(ii) Contribution to the Project: In stage 1 of each round, each group member is endowed with 20 tokens. You have to decide how many of the 20 tokens you are going to contribute to the project. The remaining tokens will be kept in your private account.

Calculation of your payoff in stage 1: Your payoff in stage 1 consists of two components:

- **tokens you have kept** = endowment – your contribution to the project
- **earnings from the project** = 1.6 x sum of the contributions of all group members / number of group members

Thus, **your payoff in Stage 1** amounts to:
20 – your contribution to the project
+ 1.6 x sum of the contributions of all group members / number of group members

The earnings from the project are calculated according to this formula for each group member. **Please note:** Each group member receives the same earnings from the project, i.e. each group member benefits from **all** contributions to the project.

Stage 2

Assignment of Tokens: In stage 2 it will be displayed how many tokens each group member contributed to the project. (**Please note: In each round the order of displaying the members' actions will randomly be determined.** Thus, it is not possible to identify any group member by her position on the displayed list throughout different rounds.) By the assignment of tokens in stage 2 you can increase or reduce the payoff of a group member or keep it unchanged.
In each round each participant receives additional 20 tokens in stage 2. You have to decide how many of these 20 tokens you are going to assign to other group members. The remaining tokens are kept in your private account. You can check the costs of your token assignment by pressing the button "Calculation of Tokens".

- Each **positive token** that you assign to a group member **increases her payoff by 1 token**.
- Each **negative token** that you assign to a group member **reduces her payoff by 3 tokens**.
- If you assign **0 tokens** to a group member her **payoff won't change**.

Calculation of your payoff in stage 2: Your payoff in stage 2 consists of three components:

- **tokens you kept in your private account** = 20 – sum of the tokens that you have assigned to the other group members
- **increased by the number of positive tokens** you have received from other group members
- **diminished by the threefold number of negative tokens** you have received from other group members

Thus, **your payoff in Stage 2** amounts to:
20 – sum of the tokens that you assigned to other group members
+ the number of positive tokens you received from other group members
– 3x (the number of negative tokens you received from other group members)

Calculation of your round payoff: Your round payoff is composed of

	Your payoff from Stage 1	20 – your contribution to the project + 1.6 x sum of the contributions of all group members / number of group members
+	Your payoff from Stage 2	20 – sum of the tokens that you have assigned to other group members + number of positive tokens you have received from other members – 3 x (the number of tokens you have received from other members)
=	Your round payoff	

Special case: a single group member: If it happens that you are the only member in your group you receive 20 tokens in Stage 1 and 20 tokens in Stage 2, i.e. your round payoff sums up to 40. You do not have to take any action neither on Stage 1 nor on Stage 2.

Information at the end of the round: At the end of the round you receive a detailed overview of the results obtained in all groups. For every group member you are informed about her: Contribution to the project, payoff from the Stage 1, assigned tokens (if possible), received positive tokens (if possible), received negative tokens (if possible), payoff from Stage 2, round payoff.

History: Starting from the 2nd round, in the beginning of a new round you receive an overview of the average results (as above) of all previous rounds.

Total Payoff: The total payoff from the experiment is composed of the initial endowment of 1,000 tokens plus the sum of round payoffs from all 30 rounds. At the end of the experiment your total payoff will be converted into Euro with an exchange rate of 1 € per 100 tokens.

Please notice: Communication is not allowed during the whole experiment. If you have a question please raise your hand out of the cabin. All decisions are made anonymously, i.e. no other participant is informed about the identity of someone who made a certain decision. The payment is anonymous too, i.e. no participant learns what the payoff of another participant is.

We wish you success!

Appendix 2.1. Instructions to the experiment PUN Treatment

General Information: At the beginning of the experiment you will be randomly assigned to one of **2 subpopulations each consisting of 12 participants**. During the whole experiment you will interact only with the members of your subpopulation. At the beginning of the experiment, **1000 experimental tokens** will be assigned to the experimental account of each participant.

Course of Action: The experiment consists of **30 rounds**. Each round consists of 2 stages. In Stage 1, the group choice and the decision regarding the contribution to the project take place. In Stage 2, participants may influence the earnings of the other group members.

Stage 1

(i) The Group Choice: In Stage 1, each participant decides which group she wants to join. There are two different groups that can be joined:

	Influence on the earnings of other group members
Group	A: No
	B: Yes, by assigning negative points

(ii) Contributing to the Project: In stage 1 of each round, each group member is endowed with 20 tokens. You have to decide how many of the 20 tokens you are going to contribute to the project. The remaining tokens will be kept by you.

Calculation of your payoff in stage 1: Your payoff in stage 1 consists of two components:

- **tokens you have kept** = endowment – your contribution to the project
- **earnings from the project** = 1.6 x sum of the contributions of all group members / number of group members

Thus, **your payoff in Stage 1** amounts to:

20 – your contribution to the project
+ 1.6x sum of the contributions of all group members / number of group members

The earnings from the project are calculated according to this formula for each group member. **Please note:** Each group member receives the same earnings from the project, i.e. each group member benefits from **all** contributions to the project.

Stage 2 - Assignment of Tokens: In stage 2 it will be displayed how much each group member contributed to the project. (**Please note: Before each round a display order will randomly be determined.** Thus, it is not possible to identify any group member by her position on the displayed list throughout different rounds.) By the assignment of tokens you can reduce the payoff of a group member or keep it unchanged.

In each round each participant receives additional 20 tokens in stage 2. You have to decide how many from the 20 tokens you are going to assign to other group members. The remaining

tokens are kept by yourself. You can check the costs of your token assignment by pressing the button *Calculation of Tokens*.

- Each **negative token** that you assign to a group member **reduces her payoff by 3 tokens**.
- If you assign **0 tokens** to a group member her **payoff won't change**.

Calculation of your payoff in stage 2: Your payoff in stage 2 consists of two components:

- **tokens you kept** = 20 – sum of the tokens that you have assigned to the other group members
- **less the threefold number of negative tokens** you have received from other group members

Thus, **your payoff in Stage 2** amounts to:
20 – sum of the tokens that you assigned to other group members

– 3x (the number of tokens you received from other group members)

Calculation of your round payoff: Your round payoff is composed of

	Your payoff from Stage 1	20 – your contribution to the project + 1.6 x sum of the contributions of all group members / number of group members
+	Your payoff from Stage 2	20 – sum of the tokens that you have assigned to other group members – 3 x (the number of tokens you have received from other group members
=	Your round payoff	

Special case: a single group member: If it happens that you are the only member in your group you receive 20 tokens in Stage 1 and 20 tokens in Stage 2, i.e. your round payoff amounts to 40. You neither have to take any action on Stage 1 nor on Stage 2.

Information at the end of the round: At the end of the round you receive a detailed overview of the results obtained in all groups. For every group member you are informed about her: Contribution to the project, payoff from the Stage 1, assigned tokens (if possible), received tokens (if possible), payoff from Stage 2, round payoff.

History: Starting from the 2nd round, in the beginning of a new round you receive an overview of the average results (as above) of all previous rounds.

Total Payoff: The total payoff from the experiment is composed of the starting capital of 1000 tokens plus the sum of round payoffs from all 30 rounds. At the end of the experiment your total payoff will be converted into Euro with an exchange rate of 1 € per 100 tokens.

Please notice: Communication is not allowed during the whole experiment. If you have a question please raise your hand out of the cabin. All decisions are made anonymously, i.e. no other participant is informed about the identity of someone who made a certain decision. The payment is anonymous too, i.e. no participant learns what the payoff of another participant is.

Appendix 2.2. Corollaries, proofs and calculations

Corollary 1 (to Proposition 4 (b) in FS) From Proposition 4 (b) of FS (p. 849) we know that if the number of free-riders k_N is sufficiently high, i.e., if

(A1) $k_N(n_N - 1) \geq a/2$,

then there is a unique equilibrium in which no player contributes. In our model, for a_N, we simply insert R/n_N in (A1) and obtain $k_N \geq \frac{n_N - 1}{2n_N} R$. This means for our setting that if $k_N \geq \frac{n_N - 1}{2n_N} R$ then there exists a unique equilibrium in NSC in which all players contribute $g_N^* = 0$ to the joint project. Note that since $0 \leq \frac{n_N - 1}{2n_N} < \frac{1}{2}$, $R/2$ is an upper bound for $\frac{n_N - 1}{2n_N} R$. Hence, if the number of free-riders exceeds $R/2$ we can be sure that no equilibrium with strictly positive contributions exists.

Corollary 2 (to Proposition 4 in FS) If all players in NSC are conditional cooperators then there exist equilibria in NSC in which all players contribute $g_N \in \{0...y\}$ to the joint project.

Proof to Corollary 2 A conditional cooperator is ready to increase his/her own contribution if all others contribute more. Suppose all players are conditional cooperators and contribute g_N. In this case a conditional cooperator has no incentive to deviate to a lower contribution $g_i < g_N$ since for him or her (by assumption) the monetary benefit of withholding a dollar is lower than the total loss from deviation $a_\theta + \beta$.

Proof to Lemma 1 Without any punishment, the players who contribute g have a lower monetary payoff than the deviator who contributes $g_i < g$ and in addition a disutility from inequality to their disadvantage. Assume that each of the n_S' players punishes with the same punishment amount γ^*. Then, the monetary payoff of a punisher j who contributes g and punishes with γ^* is: $x_j = y - g + a_S[(n_S - 1)g + g_i] - c\gamma^*$ The payoff of the contribution deviator i who contributes $g_i < g$ is: $x_i = y - g_i + a_S[(n_S - 1)g + g_i] - n_S'\gamma^*$. To determine the punishment level γ^* that equalizes the monetary payoff of the $n_S' \leq n_S - 1$ punishers and the deviator, we set $x_i = x_j$ and obtain $\gamma^* = \frac{g - g_i}{n_S' - c}$. Q.E.D.

Proof to Proposition 1 Suppose that the contribution deviator i is punished with the punishment level γ^* by each of the n_S' punishers. According to Lemma 1 this leads to equal monetary payoffs of the contribution deviator and each of the n_S' punishers. However, a punisher may increase its monetary payoff by deviating to a lower punishment level $\gamma^D < \gamma^*$. At the same time, this deviation generates (a) advantageous inequality in comparison to the other punishers, (b) disadvantageous inequality to the contribution deviator and (c) it decreases the disadvantageous inequality to the contributors who do not punish. To answer

the question whether a deviation from the punishment level γ^* is profitable, we calculate the net utility change through deviation in punishment. The payoff of the punishment deviator who contributes g but punishes just γ^D is:

$$x_j^D = y - g + a_S[(n_S - 1)g + g_i] - c\gamma^D$$

The payoff of the contribution deviator i who contributes $g_i < g$ and does not punish is:

$$x_i = y - g_i + a_S[(n_S - 1)g + g_i] - (n_S'-1)\gamma^* - \gamma^D$$

Next to this player there are $(n_S - n_S'-1)$ non-punishing contributors l who contribute g but refrain from the punishment. The payoff of each of these players is:

$$x_l = y - g + a_S[(n_S - 1)g + g_i]$$

The payoff of each of the $(n_S'-1)$ punishers m who contribute g and punish with γ^* is:

$$x_m = y - g + a_S[(n_S - 1)g + g_i] - c\gamma^*$$

According to (4) the utility of the punishment deviator is then

$$\begin{aligned}(A2)\quad u_j^D &= y - g + a_S[(n_S - 1)g + g_i] - c\gamma^D \\ &- \left(\frac{\alpha_j}{n_S - 1}[g - g_i + c\gamma^D - (n_S'-1)\gamma^* - \gamma^D]\right) - \left(\frac{\alpha_j}{n_S - 1}(n_S - n_S'-1)c\gamma^D\right) \\ &- \left(\frac{\beta_j}{n_S - 1}(n_S'-1)(\gamma^* - \gamma^D)c\right).\end{aligned}$$

In the first line of (A2) the monetary payoff of the punishment deviator is shown; the first term in the second line (in big parentheses) shows the utility loss through disadvantageous inequality with respect to the contribution deviator who contributes $g_i < g$. The second term in the second line reflects the utility loss due to the payoff inequity between the punishment deviator and the contributors who do not punish. Finally, the term in the third line shows the utility loss due to the inequality in payoffs between the punishment deviator and players who punish with level γ^*.

Whether the utility from deviating to a punishment level γ^D exceeds the utility from sticking to γ^*, depends on the costs of punishment in relation to its "gain". The utility of the deviator (A2) is increasing in γ^D until $\gamma^D = \gamma^*$, i.e., $\frac{\partial u_j^D}{\partial \gamma^D} > 0$. As shown below in the calculations $\frac{\partial u_j^D}{\partial \gamma^D} > 0$ if and only if $c < \frac{\alpha_j}{(n_S - 1)(1 + \alpha_j) - (n_S'-1)(\alpha_j + \beta_j)}$. Q.E.D.

Calculations to Proposition 1 The derivative of (A2) with respect to γ^D is:

(A3) $$\frac{\partial u_j^D}{\partial \gamma^D} = -c - c\frac{\alpha_j}{n_S - 1} + \frac{\alpha_j}{n_S - 1} - \frac{\alpha_j}{n_S - 1}(n_S - n_S' - 1)c + \frac{\beta_j}{n_S - 1}(n_S' - 1)c$$

Since $\frac{\partial u_j^D}{\partial \gamma^D}$ is independent of γ^D, u_j^D is maximized with $\gamma^D = \gamma^*$ if the right-hand side of (A3) is strictly positive.

$$-c - c\frac{\alpha_j}{n_S - 1} + \frac{\alpha_j}{n_S - 1} - \frac{\alpha_j}{n_S - 1}(n_S - n_S' - 1)c + \frac{\beta_j}{n_S - 1}(n_S' - 1)c > 0$$

$$\Leftrightarrow 1 + \frac{\alpha_j}{n_S - 1} - \frac{\alpha_j}{(n_S - 1)c} + \frac{\alpha_j}{n_S - 1}(n_S - n_S' - 1) - \frac{\beta_j}{n_S - 1}(n_S' - 1) < 0$$

$$\Leftrightarrow (n_S - 1) + \alpha_j - \frac{\alpha_j}{c} + \alpha_j(n_S - n_S' - 1) - \beta_j(n_S' - 1) < 0$$

$$\Leftrightarrow (n_S - 1) + \alpha_j + \alpha_j(n_S - n_S' - 1) - \beta_j(n_S' - 1) < \frac{\alpha_j}{c}$$

Solving for *c* yields condition (7), i.e., $c < \frac{\alpha_j}{(n_S - 1)(1 + \alpha_j) - (n_S' - 1)(\alpha_j + \beta_j)}$.

Proof to Proposition 2 In Corollary 2 we have already shown that even without the threat of punishment, no conditional cooperator has an incentive to deviate if all others contribute *g*, because the utility from the material gain from a deviation would be (more than) destroyed by the utility loss from advantageous inequality. Here we show that no player (neither a conditional cooperator nor a free-rider) has an incentive to deviate if (7) is satisfied. Suppose that player i deviates by contributing $g_i < g$. Then (as shown in Proposition 1) the enforcers will punish her with $\gamma^* = \frac{g - g_i}{n_S' - c}$. The utility of the player i then is:

(A4) $$u_i = y - g_i + a_S[(n_S - 1)g + g_i] - \frac{\beta_i}{n_S - 1}(n_S - 1)(g - g_i) - n_S'\gamma^*$$

The derivative of (A4) with respect to g_i is $\frac{\partial u_i}{\partial g_i} = a_S + \beta_i - 1 + \frac{n_S'}{n_S' - c}$. Since $\frac{n_S'}{n_S' - c} > 1 - (a_S + \beta_i)$ for all $c < 1$ it follows that $g_i = g$ maximizes the utility. Thus, a deviation to $g_i < g$ is not profitable. Q.E.D.

Proof to Proposition 3 If no player satisfies condition (7) then punishment is not credible and absent. We already have shown in Corollary 1 that in a non-sanctioning-community cooperation is not possible if the number of free-riders exceeds $\frac{n_S - 1}{2n_S}R$. Q.E.D.

Proof to Corollary 3 Assume the existence of n_S' enforcers. Each of the n_S' enforcers *j* has to satisfy (7), which can be rephrased as

(A5) $c(n_S - n_S') < 1 - \frac{c(n_S - 1) + c\beta_j(n_S' - 1)}{\alpha_j}$.

By contradiction we show that $c(n_S - n_S') < 1$ has to be satisfied. Suppose $c(n_S - n_S') \geq 1$. Then (A5) would imply that $1 - \frac{c(n_S - 1) + c\beta_j(n_S' - 1)}{\alpha_j} > 1$ which is equivalent to $\frac{c(n_S - 1) + c\beta_j(n_S' - 1)}{\alpha_j} < 0$. This, however, can never be the case because $\frac{c(n_S - 1) + c\beta_j(n_S' - 1)}{\alpha_j}$ is always strictly positive for each of the n_S' enforcers. Hence $c(n_S - n_S') < 1$ has to be satisfied. Q.E.D.

Proof to Lemma 2 Assume that each of the n_S' players rewards every other contributor with the same reward amount ρ^S. The costs of rewarding are assumed to be strictly lower than 1, i.e., $c < 1$. Then, the payoff of a reward provider j who contributes g and rewards with ρ^S is: $x_j = y - g + a_S n_S' g - (n_S' - 1)\rho^S c + (n_S' - 1)\rho^S$. The payoff of a free-rider i who refrains from contributing and rewarding is: $x_i = y + a_S n_S' g$. To determine the reward level ρ^S that equalizes the monetary payoff of a reward provider and a free-rider, we set $x_i = x_j$ and solve for ρ^S. We obtain $\rho^S = \frac{g}{(1-c)(n_S' - 1)}$. Rewarding with $\rho < \rho^S$ does clearly not pay since in this case contributors suffer from disadvantageous inequality with respect to the free-riders. Q.E.D.

Proof to Proposition 4 We have to check whether one of the $n_S - n_S'$ free-riders has an incentive to deviate and contribute also $g > 0$ in order to be rewarded by the conditional cooperators. If deviation is profitable then there exists no separating equilibrium with the properties described in Proposition 4. The payoff of the deviating free-rider who contributes $g > 0$ but refrains from rewarding is:

$$x_i^D = y - g + a_S(n_S' + 1)g + n_S'\rho^S$$

The payoff of each of the $(n_S - n_S' - 1)$ free-riders who contribute zero is:

$$x_i = y + a_S(n_S' + 1)g$$

The payoff of each of the n_S' reward providers who contribute $g > 0$ and reward with ρ^S is:

$$x_j = y - g + a_S(n_S' + 1)g + (n_S' - 1)\rho^S - n_S'\rho^S c$$

The utility of the deviating free-rider then is:

(A6) $$u_i^D = y - g + a_S(n_S' + 1)g + n_S'\rho^S$$
$$- \left(\frac{\beta_i}{n_S - 1}(n_S - n_S' - 1)(n_S'\rho^S - g) \right) - \left(\frac{\beta_i}{n_S - 1} n_S'\rho^S(1 + n_S' c) \right)$$

In the first line of (A6) the monetary payoff of the deviating free-rider is shown; the term in the second line (in big parentheses) shows the utility loss through advantageous inequality with respect to the $(n_S - n_S'-1)$ free-riders who do not contribute. The term in the third line reflects the utility loss due to the advantageous inequality with respect to the n_S' reward providers who reward with level ρ^S.

To check whether a free-rider has an incentive to deviate, we have to compare the player's utility from deviation (A6) with the utility the player would obtain from the equilibrium strategy. Since in equilibrium all players would obtain the same monetary payoff the utility of a free-rider is (who contributes zero in equilibrium) is:

$$u_i = y + a_S n_S' g$$

Hence, if $u_i^D - u_i > 0$, i.e., if

$$u_i^D - u_i = y - g + a_S(n_S'+1)g + n_S'\rho^S - \frac{\beta_i}{n_S - 1}(n_S - n_S'-1)(n_S'\rho^S - g)$$

$$-\frac{\beta_i}{n_S - 1} n_S'\rho^S(1 + n_S'c) - y - a_S n_S' g > 0$$

then the player deviates. We insert the payoff equalizing reward level in the inequality (with n_S' reward providers and $n_S'+1$ reward recipients) $\rho^S = \frac{g}{(1-c)n_S'}$ and solve for β_i:

$$\text{(A7)} \quad \beta_i < \frac{(n_S - 1)[a_S + c(1 - a_S)]}{(n_S - 1)c + 1}$$

If a free-rider has a sufficiently low β_i then this player deviates from contributing zero to contributing g in order to be rewarded on the second stage. However, according to the assumption of the FS model a free-rider satisfies the condition $\beta_i < 1 - a_S$. It can be easily shown that if the productivity of the joint project is strictly higher than 1 then the right side of the condition (A7) is always greater than $1 - a_S$, i.e., if $R > 1$ then $1 - a_S < \frac{(n_S - 1)[a_S + c(1 - a_S)]}{(n_S - 1)c + 1}$. This means, a free-rider satisfies always the condition (A7) thus having always an incentive to deviate. Hence, a separating equilibrium as described in Proposition 4 cannot exist.

Proof to Proposition 5 Suppose that each player who contributes $g > 0$ in Stage 1 (i.e., all players) is rewarded with ρ^* by each of the n_S' reward providers. However, a reward provider may increase monetary payoff by deviating from the reward level ρ^* to a lower reward level ρ^D. By deviating to $\rho^D < \rho^*$, the reward deviator increases own payoff and creates a payoff inequality to the own advantage with respect to other reward providers who stick to ρ^*. On the other hand, the reward deviator suffers from inequality to the own disadvantage with respect to the reward free-riders. To answer the question whether a deviation from the reward level ρ^* is profitable, we calculate the net utility change through

the deviation in rewarding. The payoff of the reward deviator who contributes $g > 0$ but rewards just ρ^D is:

$$x_j^D = y - g + a_S n_S g - (n_S - 1)c\rho^D + (n_S' - 1)\rho^*$$

The payoff of each of the $(n_S - n_S')$ reward free-riders who contribute $g > 0$ but do not reward is:

$$x_i = y - g + a_S n_S g + (n_S' - 1)\rho^* + \rho^D$$

If one of the reward providers deviates then there are $(n_S' - 1)$ reward providers who contribute $g > 0$ and reward with ρ^*. The payoff of each of these players is:

$$x_j = y - g + a_S n_S g - (n_S - 1)c\rho^* + (n_S' - 2)\rho^* + \rho^D$$

According to (4), the utility of the reward deviator is then:

$$\text{(A8)} \quad u_j^D = y - g + a_S n_S g - (n_S - 1)c\rho^D + (n_S' - 1)\rho^* - \left(\frac{\alpha_j}{n_S - 1}(n_S - n_S')\rho^D[1 + (n_S - 1)c]\right) - \left(\frac{\beta_j}{n_S - 1}(n_S' - 1)(\rho^* - \rho^D)[1 + (n_S - 1)c]\right)$$

In the first line of (A8) the monetary payoff of the reward deviator is shown; the term in the second line (in big parentheses) shows the utility loss through disadvantageous inequality with respect to the $(n_S - n_S')$ free-riders. The term in the third line reflects the utility loss due to advantageous inequality with respect to the other $(n_S' - 1)$ reward providers who reward with level ρ^*. Whether the utility from deviating to a reward level ρ^D exceeds the utility from sticking to reward level ρ^* depends on the costs of rewarding in relation to its "gain". The utility of the reward deviator (A8) is increasing in ρ^D until $\rho^D = \rho^*$ if $\frac{\partial u_j^D}{\partial \rho^D} > 0$. As shown below in the calculations $\frac{\partial u_j^D}{\partial \rho^D} > 0$ if and only if $c < \frac{\beta_j(n_S' - 1) - \alpha_j(n_S - n_S')}{(n_S - 1)[n_S - 1 - \beta_j(n_S' - 1) + \alpha_j(n_S - n_S')]}$. Q.E.D.

Calculations to Proposition 5 The derivative of (A8) with respect to ρ^D is:

$$\text{(A9)} \quad \frac{\partial u_j^D}{\partial \rho^D} = -(n_S - 1)c - \frac{\alpha_j}{n_S - 1}(n_S - n_S')[1 + (n_S - 1)c] + \frac{\beta_j}{n_S - 1}(n_S' - 1)[1 + (n_S - 1)c]$$

Since $\frac{\partial u_j^D}{\partial \rho^D}$ is independent of ρ^D, u_j^D is maximized with $\rho^D = \rho^*$ if the right-hand side of (A9) is strictly positive. It can be easily seen that it never pays to reward by more than ρ^*.

$$-(n_S-1)c-\frac{\alpha_j}{n_S-1}(n_S-n_S')-\alpha_j(n_S-n_S')c+\frac{\beta_j}{n_S-1}(n_S'-1)+\beta_j(n_S'-1)c>0$$

$$\Leftrightarrow[-(n_S-1)-\alpha_j(n_S-n_S')+\beta_j(n_S'-1)]c>\frac{1}{n_S-1}[\alpha_j(n_S-n_S')-\beta_j(n_S'-1)]$$

$$\Leftrightarrow[(n_S-1)+\alpha_j(n_S-n_S')-\beta_j(n_S'-1)]c<\frac{1}{n_S-1}[\beta_j(n_S'-1)-\alpha_j(n_S-n_S')]$$

Solving for *c* yields condition (8), i.e., $c<\frac{\beta_j(n_S'-1)-\alpha_j(n_S-n_S')}{(n_S-1)[n_S-1-\beta_j(n_S'-1)+\alpha_j(n_S-n_S')]}$.

Appendix 3.1. Instructions to the experiment

General Information: At the beginning of the experiment you will be randomly assigned to one of **2 subpopulations each consisting of 12 participants**. During the whole experiment you will interact only with the members of your subpopulation. At the beginning of the experiment, **1000 experimental tokens** will be assigned to the experimental account of each participant.

Course of Action: The experiment consists of **30 rounds**. Each round consists of 2 stages. In Stage 1, the group choice and the decision regarding the contribution to the project take place. In Stage 2, participants may influence the earnings of the other group members.

Stage 1

(i) The Group Choice: In Stage 1, each participant decides which group she wants to join. There are two different groups that can be joined:

	Influence on the earnings of other group members
Group	A: No
	B: Yes, by assigning negative points

(ii) Contributing to the Project: In stage 1 of each round, each group member is endowed with 20 tokens. You have to decide how many of the 20 tokens you are going to contribute to the project. The remaining tokens will be kept by yourself.

Calculation of your payoff in stage 1: Your payoff in stage 1 consists of two components:

- **tokens you have kept** = endowment – your contribution to the project
- **earnings from the project** = 1.6 x sum of the contributions of all group members / number of group members

Thus, **your payoff in Stage 1** amounts to:
20 – your contribution to the project
+ 1.6x sum of the contributions of all group members / number of group members

The earnings from the project are calculated according to this formula for each group member. **Please note:** Each group member receives the same earnings from the project, i.e. each group member benefits from **all** contributions to the project.

Stage 2: Assignment of Tokens: In stage 2 it will be displayed how much each group member contributed to the project. (**Please note: Before each round a display order will randomly be determined.** Thus, it is not possible to identify any group member by her position on the displayed list throughout different rounds.) By the assignment of tokens you can reduce the payoff of a group member or keep it unchanged.

In each round each participant receives additional 20 tokens in stage 2. You have to decide how many from the 20 tokens you are going to assign to other group members. The remaining tokens are kept by yourself. You can check the costs of your token assignment by pressing the button *Calculation of Tokens*.

- Each **negative token** that you assign to a group member **reduces her payoff by 3 tokens**.
- If you assign **0 tokens** to a group member her **payoff won't change**.

Calculation of your payoff in stage 2: Your payoff in stage 2 consists of two components:

- **tokens you kept** = 20 – sum of the tokens that you have assigned to the other group members

- **less the threefold number of negative tokens** you have received from other group members

Thus, **your payoff in Stage 2** amounts to:
20 – sum of the tokens that you assigned to other group members
– 3x (the number of tokens you received from other group members)

Calculation of your round payoff: Your round payoff is composed of

	Your payoff from Stage 1	20 – your contribution to the project + 1.6 x sum of the contributions of all group members / number of group members
+	Your payoff from Stage 2	20 – sum of the tokens that you have assigned to other group members – 3 x (the number of tokens you have received from other group members
=	Your round payoff	

Special case: a single group member: If it happens that you are the only member in your group you receive 20 tokens in Stage 1 and 20 tokens in Stage 2, i.e., your round payoff amounts to 40. You neither have to take any action on Stage 1 nor on Stage 2.

Information at the end of the round: At the end of the round you receive a detailed overview of the results obtained in all groups. For every group member you are informed about her: Contribution to the project, payoff from the Stage 1, assigned tokens (if possible), received tokens (if possible), payoff from Stage 2, round payoff.

History: Starting from the 2nd round, in the beginning of a new round you receive an overview of the average results (as above) of all previous rounds.

Report sheet about the decisions of participants of a previously conducted experiment

Each participant receives a report sheet about the decisions of participants of a previous experiment which was conducted in the eLab at the University Erfurt in January 2004. In this report you will find average numbers of the decisions of the participants. Please read this report before you decide.

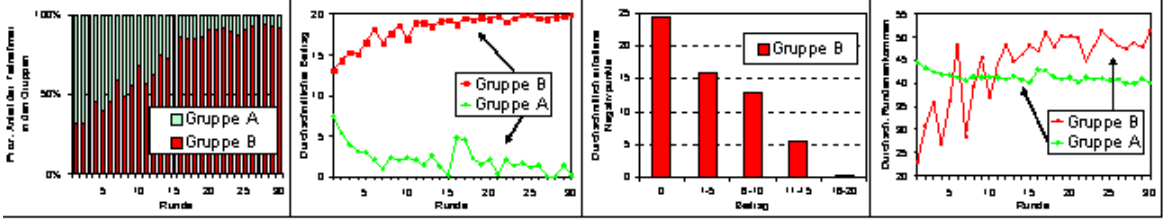
Total Payoff: The total payoff from the experiment is composed of the starting capital of 1000 tokens plus the sum of round payoffs from all 30 rounds. At the end of the experiment your total payoff will be converted into Euro with an exchange rate of 1 € per 100 tokens.

Please notice: Communication is not allowed during the whole experiment. If you have a question please raise your hand out of the cabin. All decisions are made anonymously, i.e. no other participant is informed about the identity of someone who made a certain decision. The payment is anonymous too, i.e. no participant learns what the payoff of another participant is. We wish you success!

Appendix 3.2. Report sheet for the experiment

Report sheet about the participants' decisions of a previous experiment conducted in January 2004

Average number of group members		Average contribution		Average received negative tokens		Average payoff	
A	**B**	**A**	**B**	**A**	**B**	**A**	**B**
8.3	3.8	7.4	13.1	-	6.2	44.4	23.0
8.3	3.8	5.4	14.4	-	4.5	43.2	30.8
8.3	3.8	3.9	15.3	-	3.3	42.4	35.8
6.5	5.5	3.2	15.1	-	5.6	41.9	26.7
7.3	4.8	2.9	16.6	-	3.5	41.7	36.1
6.6	5.4	2.1	18.3	-	0.7	41.3	48.3
5.0	7.0	0.9	16.5	-	5.3	40.6	28.6
6.1	5.9	2.4	17.7	-	2.9	41.4	39.2
5.4	6.6	2.1	18.6	-	1.5	41.2	45.4
3.9	8.1	2.3	17.0	-	3.3	41.4	36.8
5.3	6.8	2.1	19.0	-	1.8	41.3	44.3
4.5	7.5	1.4	19.0	-	0.8	40.8	48.1
3.0	9.0	2.6	18.5	-	1.6	41.5	44.8
3.4	8.6	1.2	19.2	-	1.3	40.7	46.1
2.4	9.6	0.2	19.3	-	0.9	40.1	48.1
1.8	10.3	4.8	18.8	-	1.1	42.9	46.9
1.9	10.1	4.6	19.5	-	0.2	42.8	50.9
1.9	10.1	2.2	19.3	-	0.9	41.3	47.9
1.6	10.4	1.5	19.6	-	0.4	40.9	50.2
1.1	10.9	2.1	19.4	-	0.4	41.3	50.2
1.1	10.9	0.3	19.7	-	0.5	40.2	49.9
1.0	11.0	2.1	19.1	-	1.7	41.3	44.6
1.3	10.8	1.3	19.5	-	0.9	40.8	48.0
1.5	10.5	1.7	19.9	-	0.2	41.0	51.3
1.1	10.9	1.1	19.9	-	0.6	40.7	49.6
0.9	11.1	1.4	19.6	-	0.9	40.9	48.1
0.5	11.5	0.0	19.5	-	1.1	40.0	47.5
0.8	11.3	0.0	19.7	-	0.7	40.0	48.8
0.9	11.1	1.4	19.8	-	1.0	40.9	47.7
1.0	11.0	0.1	20.0	-	0.2	40.1	51.3
Average over all periods							
3.4	**8.6**	**2.9**	**18.8**	-	**1.4**	**41.8**	**45.7**



Appendix 4.1. Instructions to the experiment

General Information: At the beginning of the experiment you are randomly assigned to groups each consisting of **6 participants**. One of the members in your group is randomly chosen as Type A participant. All other members in your group become Type B participants. During the whole experiment your type does not change and you only interact with the members of your own group. At the beginning of the experiment, **500 experimental tokens** are assigned to the experimental account of each participant.

Course of Action: The experiment consists of **30 rounds** containing three blocks of periods of 10 rounds each. Before each block starts (i.e., before Round 1, 11, and 21) the Type A player chooses between two modes of token allocation: "**allocation of positive tokens**" or "**allocation of negative tokens**".

Each round consists of 2 stages. In stage 1, each group member (Type A as well as Type B participant) decides on the individual contribution to the project. In stage 2 Type A participant may influence the earnings of the other group members by allocating tokens.

Stage 1: Contributing to the Project: In stage 1 of each round, each group member is endowed with 20 tokens. You have to decide how many of the 20 tokens you are going to contribute to the project. The remaining tokens are kept in your account.

Calculation of your payoff in stage 1: Your payoff in stage 1 consists of two components:

- **tokens you keep** = endowment – your contribution to the project
- **earnings from the project** = 1.6 x sum of the contributions of all group members / number of group members

Thus, **your payoff in stage 1** amounts to:
20 – your contribution to the project
+ 1.6 x sum of the contributions of all group members / number of group members

The earnings from the project are calculated according to this formula for each group member. **Please note:** Each group member receives the same earnings from the project, i.e. each group member benefits from **all** contributions to the project.

Stage 2: Assignment of Tokens by the Type A participant: In stage 2 the Type A participant gets informed about how many tokens each group member contributed to the project. (**Please note: Before each round the display order of the Type B participants is randomly determined.** Thus, it is not possible to identify a Type B participant by his or her position on the displayed list throughout different rounds.)

By the assignment of tokens the Type A participant can increase or reduce the payoff (according to the chosen mode for the allocation of tokens) of a group member or keep it unchanged.

In each round the type A participant receives additional 20 tokens in stage 2. The type A participant has to decide how many of the 20 tokens he or she is going to assign to each of the other group members. The remaining tokens are kept by the Type A participant. The Type A participant can check the costs of the token assignment by pressing the button *Calculation of Tokens*.

- The assignment of **0 tokens** to a Type B participant **does not change his or her payoff**.
- If the mode "**allocation of positive tokens**" is chosen for the current 10 block period, each **positive token** that is assigned to a Type B participant **increases his or her payoff by 3 tokens**.
- If the mode "**allocation of negative tokens**" is chosen for the 10 block period, each

negative token that is assigned to a Type B participant **reduces his or her payoff by 3 tokens**.

Calculation of payoffs in stage 2:

Type A participant: Your payoff in stage 2 consists of **tokens you keep**
= 20 – sum of the tokens that you have assigned to the other group members

Type B participant: If the mode "**allocation of negative tokens**" is chosen
Your (negative) payoff from stage 2 is given by 3 x the number of tokens that you have received from the Type A participant.

Calculation of your round payoff:

Your round payoff = Your payoff from stage 1 + your payoff from stage 2

Information at the end of the round: At the end of the round you receive a detailed overview of the results obtained in all groups. For every group member you are informed about his or her contribution to the project, payoff from stage 1, assigned tokens (Type A participant), received tokens (Type B participants), payoff from stage 2, round payoff. (**Please note: Before each round the display order of the Type B participants is randomly be determined.** Thus, it is not possible to identify a Type B participant by his or her position on the displayed list throughout different rounds.)

History: Starting from the 2nd round, in the beginning of a new round you receive an overview of the average results (as above) of all previous rounds. Additionally, the Type A participant receives this overview each time before each block starts i.e., when he or she decides on the mode of token allocation.

Total Payoff: The total payoff from the experiment is composed of the starting capital of 500 tokens plus the sum of round payoffs from all 30 rounds. At the end of the experiment your total payoff will be converted into Euro with an exchange rate of 1 € per 100 tokens.

Please notice: Communication is not allowed during the whole experiment. If you have any questions please raise your hand. All decisions are made anonymously, i.e. no other participant is informed about the identity of someone who made a certain decision. The payment is anonymous too, i.e. no participant gets to know the payoff of another participant.

We wish you success!

Appendix 4.2. Supporting Figures

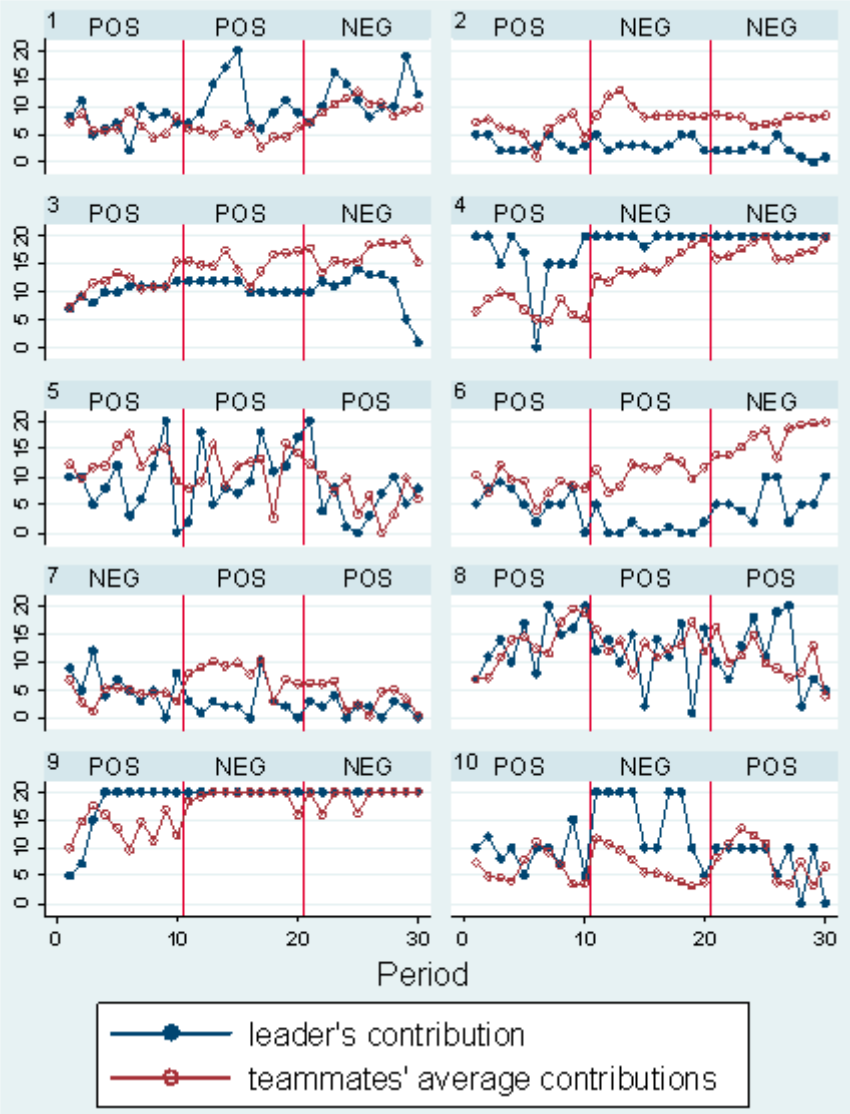
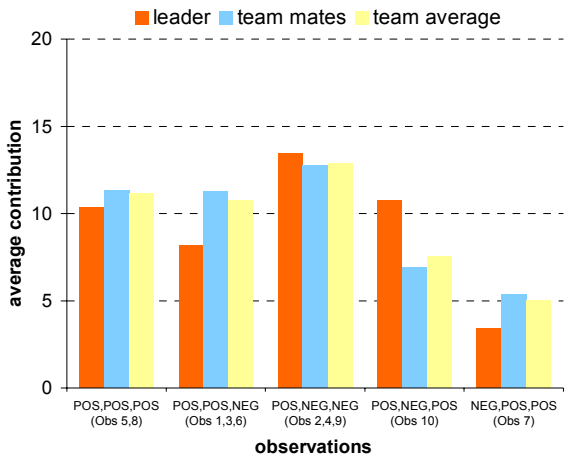


Figure A4.1 Average contributions over periods in all teams

a) incentive scheme sequences and team contributions



b) incentive scheme sequences and team payoffs

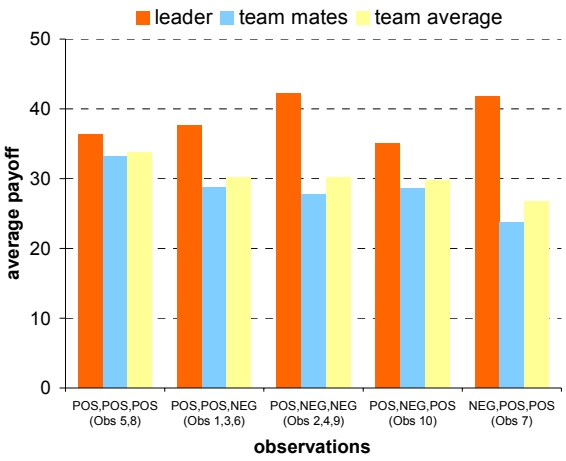


Figure A4.2 Average overall contributions and payoffs for different sequences of incentive schemes

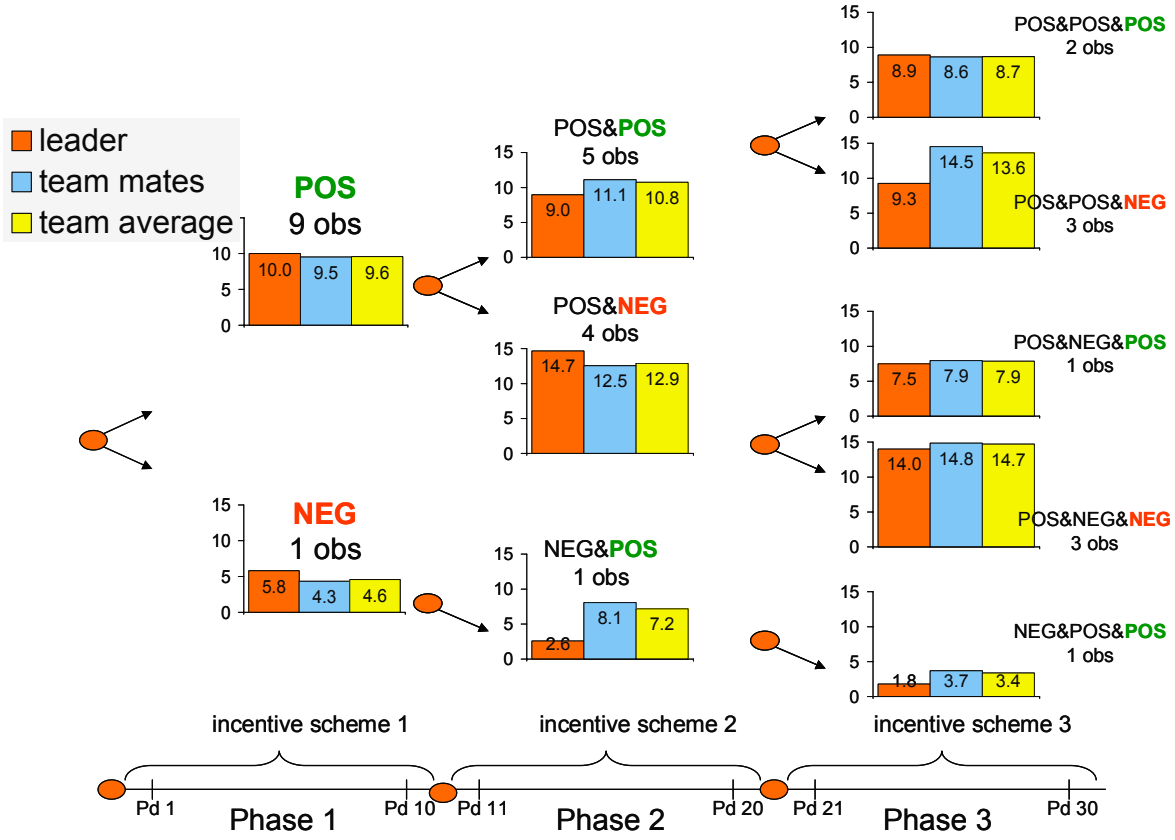


Figure A4.3: Path-dependent contributions in different phases

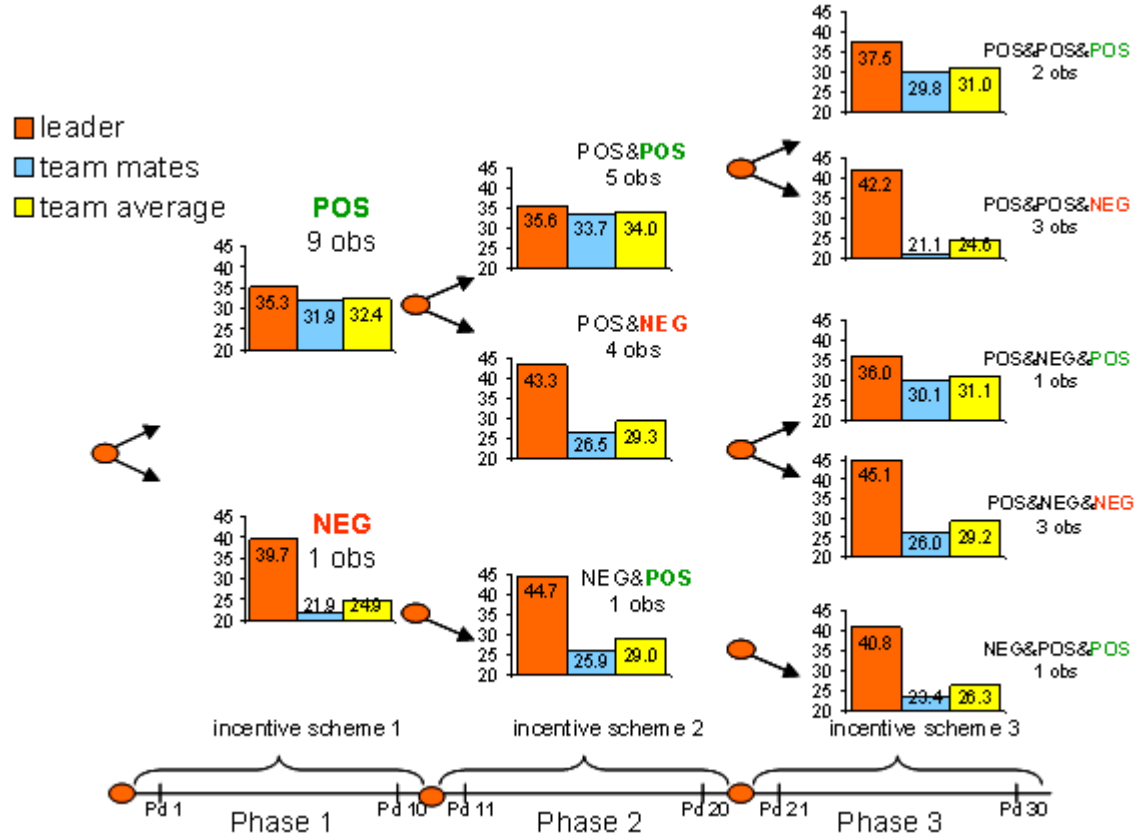


Figure A4.4: Path-dependent payoffs in different phases

Curriculum Vitae

Özgür Gürerk

Address
University of Erfurt
Microeconomics / Laboratory for Experimental Economics
Nordhäuser Str. 63
99089 Erfurt
Deutschland
Tel: +49 (0) 361 737 4561
Fax: +49 (0) 361 737 4529
e-mail: guererk@uni-erfurt.de

Current Position

since 12/2007	Research Associate at the Laboratory for Experimental Economics, University of Erfurt

Education & Previous Positions

12/2001-11/2007	Research Assistant and Ph.D. candidate at the University of Erfurt
10/1995-10/2001	Economics at the University of Bonn ("Diplom-Volkswirt") Diploma-Thesis: Der Einfluss von Gewinntabellen auf das Verhalten in Oligopolexperimenten (The Effect of Profit Tables on Behavior in Oligopoly Experiments), Supervisor: Prof. Dr. Reinhard Selten
05/1998-09/2001	Student Research Assistant at the Laboratory for Experimental Economics, University of Bonn

Research interests
Behavioral economics, microeconomics, game theory, social dilemmas