

Voting with Feet – Community Choice in Social Dilemmas

ÖZGÜR GÜRERK, BERND IRLENBUSCH, BETTINA ROCKENBACH

October 2011

Abstract

Although economic interactions often take place in open communities, the dynamics of the community choice process and its impact on cooperation are yet not well understood. In Gürer, Irlenbusch, and Rockenbach (2006) we show that a reward-and-punishment-community efficiently provides public goods in a voting with feet setting. To fully understand the determinants of this success we conduct a series of five new experiments. We show that in community choice with pure punishment possibilities cooperation is significantly higher than in settings with pure reward possibilities. We further show that the initial endogenous self-selection of subjects is an important key for the establishment and efficient maintenance of cooperation, while slow community growth is less decisive.

Keywords

Cooperation, social dilemmas, community choice, punishment, voting with feet

JEL-Classification

C72, C92, H41

Addresses

Özgür Gürek	Bernd Irlenbusch	Bettina Rockenbach
University of Erfurt	University of Cologne	University of Cologne
Lab for Experimental Economics	Department of Management	Department of Economics
Nordhaeuser Str. 63	Albertus Magnus Platz	Albertus Magnus Platz
D-99089 Erfurt	D-50923 Köln	D-50923 Köln
Germany	Germany	Germany
oezguer.guererk@uni-erfurt.de	bernd.irlenbusch@uni-koeln.de	bettina.rockenbach@uni-koeln.de
www.uni-erfurt.de/mikrooekonomie	www.codebe.uni-koeln.de	www.behavecon.uni-koeln.de

The consumer-voter may be viewed as picking that community which best satisfies his preference pattern for public goods. Charles M. Tiebout (1956, p. 418)

1. Introduction

Understanding the determinants and the extent of human cooperation is one of the most challenging questions in economics. Human cooperation in social dilemmas is particularly puzzling because the conflict between collective and individual interests creates the well-known free-rider problem (Hardin, 1968; Dawes, 1980; Ostrom, 1999; Bowles, 2004). In order to disentangle different motives for cooperation and defection, researchers have successfully studied behavior in controlled social dilemma experiments. In repeated interactions, cooperation turns out to be rarely stable and generally deteriorates to rather low levels over time (Davis and Holt, 1993; Ledyard, 1995; Croson, 1998; Ostrom, 1998; Fischbacher and Gächter, 2010). Possibilities to punish norm violators have been identified as forceful means to increase cooperation. Decentralized one-to-one punishment that reduces the income of the punished player is heavily used (Yamagishi, 1986; Ostrom et al., 1992; Fehr and Gächter 2000, 2002) and is efficiency increasing in longer interactions (Gächter et al., 2008), even if punishing incurs costs for the punisher.

A common feature of the vast majority of previous experimental studies on social dilemmas is that the interaction framework is *exogenously* imposed. In reality, however, humans often vote *with their feet* between different institutional frameworks governing the interaction with others. For example, volunteers choose to work for charities with appealing goals, organizational constitutions and cultures that encourage high contributions, and to interact with others who are attracted by the same goals, constitutions and cultures. People join different clubs, sports teams, and parties and commit to abide by their rules because they want to pursue certain activities and achieve certain goals with the help of others who join the same communities. Citizens move to different jurisdictions or even to different countries because these constitute a better fit for their preferences regarding public goods provision, ways of living, or the political system and because they prefer to live together with like-minded others who want to fulfill their citizen duties under the same system.¹ Along these lines, Tiebout

¹ For a prominent example in history when people voted with their feet for a political system, recall the large-scale migration in the 1950s, when thousands of East Germans fled to the West to benefit from and to contribute to the “industrial miracle”. To quote from Time Magazine: “In the only kind of voting that remains to the East Germans—what one British diplomat calls voting with their feet—they have chosen to flee the country at a rate which for the past three months has averaged a startling 1,000 refugees a day.” (21 Nov 1955).

(1956) suggests that individuals with similar preferences for different bundles of local public goods sort themselves into communities governed by different institutions in expectation of interacting with others who have chosen the same institution. He argues that if communities are sufficiently heterogeneous and consumer-voters are fully mobile, voting with feet generates considerable efficiency gains in public goods provision.

An attempt to study this suggested beneficial role of voting with feet in public goods provision was made in Gürer, Irlenbusch, and Rockenbach (2006). In this study the impact and the dynamics of voting with feet community choice in a public goods environment are experimentally investigated.² Participants can choose between a community with punishment *and* reward possibilities (sanctioning institution), and a community with no sanction possibilities at all. Allocating one punishment token costs 1 token for the punisher and reduces the account of the punished subject by 3 tokens. Allocating one reward token costs 1 token for the reward allocating player and increases the account of the rewarded player by 1 token. In the following we refer to this voting with feet mechanism as VF-P3R1. Despite an initial reluctance to join the sanctioning institution over time almost all subjects vote with their feet for that institution while the sanction-free institution becomes depopulated. Contributions in the sanctioning institution reach almost full levels and high efficiency is attained. Cooperation is surprisingly stable and even continues when the community grows large. These results document a clear competitive advantage of the sanctioning institution. They, however, leave open important questions concerning the determinants of the success. In this paper we report a series of five new experiments, designed to shed light on these questions.

Our first research question concerns the nature of the institutions at choice. In VF-P3R1 subjects can choose between a sanction-free institution and an institution which allows for both, punishment and rewards. Recent evidence from exogenously assigned institutions suggests that exactly the combination of punishment and rewards fosters cooperation and that a punishment mechanism alone might be less effective (see Andreoni et al., 2003; Sefton et al., 2007; or the recent overview articles by Milinski and Rockenbach (in press) and Balliet et al., 2011). From VF-P3R1, however, it appears that in voting with feet the punishment rather than the rewards option is the main driving force that leads to high cooperation levels (see

² An earlier study by Ehrhart and Keser (1999) also investigates the role of endogenous regrouping in public goods game without punishment. We discuss this study in section related literature.

Table 1 in Gülerk et al., 2006). From the data, however, it cannot unambiguously be inferred that punishment alone and not the combination fosters the high levels of cooperation. To address this issue, in the current paper we report three new experiments. Specifically, we ask whether voting with feet would be equally successful if subjects would not have the choice between the combined reward/punishment and a non-sanctioning institution. In the new experiments we study voting with feet between a pure punishment and a non-sanctioning community and additionally between a pure reward and a non-sanctioning community. When subjects have the choice between a punishment and a non-sanctioning institution (VF-P3), the experimental results are remarkably similar to the findings in VF-P3R1 with respect to the dynamics of endogenous choice and the cooperation level reached. If, however, subjects may just choose between a reward institution and the non-sanctioning institution (VF-R1) results significantly differ from the findings in VF-P3R1. Though there is an initial and continued acceptance for the reward institution cooperation deteriorates to very low levels. To make sure that this result is not due to the different leverages of punishment (1:3) and reward (1:1) we increased the leverage between allocated and received rewards from the efficiency neutral 1:1 (as in VF-P3R1) to an efficiency increasing reward mechanism of 1:3. In this third new experiment (VF-R3) one token of reward gained three tokens for the rewarded subject. VF-R3 confirms the power of reward communities to attract members but even this potentially efficiency generating institution fails to increase contributions. The synopsis of these three experiments clearly shows that the additional reward possibility is not essential for the success of the voting with feet mechanism. We therefore focus our further analyses on the success of voting with feet of the punishment institution.

Our second research question aims at disentangling the determinants of success inherent in the dynamic pattern of the punishment institution. The experimental results leave room for two non-exclusive explanations: the *self-selection* of subjects and the *slow growth* of the punishment community. The voting with feet choice allows for self-selection of the community members into the preferred institution and due to the different nature of the institutions – with and without punishment possibilities – the selection may be driven by heterogeneity of players. Particularly in the beginning, this selection process may initiate and foster a culture of high levels of cooperation in the punishment community (cf. Falk et al., in press; or Gächter and Thöni, 2005, Brekke et al., 2011). Later on, other participants might be attracted by the cooperation success in the punishment institution. The second explanation which is independent from the self-selection argument is that a group that starts small and

grows slowly can better coordinate on cooperation than a group that already starts at full size. Evidence pointing into this direction is presented by Weber (2006), who finds that a slow growth path improves coordination in a coordination game.³

To disentangle these two possible explanations for the success of the voting with feet mechanism, we furthermore ran two new control experiments. In the first control experiment, we simulate the same growth paths as they endogenously occurred in the VF-P3 experiment, but we exogenously allocate subjects to the two institutions in each period. Hence, institution allocations do not emerge from self-selection, but are exogenously imposed by the experimenter. We refer to this experiment as GX-P3 (growing groups with exogenous allocation). The comparison shows that contribution rates in GX-P3 are significantly lower than in VF-P3, indicating that self-selection of subjects plays a crucial role for the superior performance of the VF mechanism. A second control experiment eliminates the effects of an increasing growth path. Subjects are exogenously allocated into fixed-sized communities in which they remain for the entire experiment. We refer to this experiment as FX-P3 (fixed size groups with exogenous allocation). Contrasting FX-P3 with GX-P3 allows studying the impact of growing communities. The results of FX-P3 do not differ significantly from GX-P3, neither in contributions and punishment behavior nor in efficiency. This indicates that starting with a small group of subjects (and growing afterwards) does not per se foster high contributions. Instead, it seems that the group has to be composed of the “right” subjects, who initially establish a cooperative environment through high contributions and rigid punishment. In the voting with feet mechanism these subjects seem to find their way into the sanctioning community quite early. Hence, our findings suggest that the endogenous choice is key for the observed high levels of cooperation. The initial self-selection of subjects who see the potential of the punishment community and are willing to highly cooperate and rigorously punish defectors despite initial personal payoff disadvantages seems to be the main driving force in voting with feet.

In the next section, we discuss related work. Section 3 introduces our model. Section 4 deals with our experimental design and procedures. In section 5, we present the results and section 6 concludes.

³ If one assumes selfish preferences our game is not a coordination game since each participant has a dominant strategy to free-ride. If, however, one focuses on participants who have social preferences, the game might turn into a coordination game, for example, conditional cooperators only want to contribute if others contribute a similar amount. Coordination on one of several Pareto-ranked equilibria with different degrees of cooperation becomes necessary (see our analysis in Appendix B).

2. Related Literature

Recently a modest experimental literature on endogenous choice in social dilemma situations has emerged. One line of research investigates the endogenous choice of interaction partners in public goods settings⁴ while another strand focuses on institution choice through voting.

To the best of our knowledge, Ehrhart and Keser (1999) are the first to use endogenous regrouping in a public goods experiment. They allow subjects to move freely from one group to another. In each group and period a simple public goods game (without punishment) is played. The MPCR is decreasing in the group size but a contribution to the public good in a larger group yields a higher group return than in a smaller group. They find that high contributors are chased by free riders. This results in an unstable sorting of high and low contributors and over time in a declining trend of contributions. The results of Ehrhart and Keser (1999) show that the mere possibility to regroup does not lead to very high contributions though the contributions are higher than observed in a standard public goods game. Also outside the lab free mobility may cause unwanted phenomena like mass migration.

Coricelli et al. (2004) let subjects bid for the right to choose partners. As in Ehrhart and Keser (1999), in their unidirectional choice treatment free riders show a tendency to chase high contributors which leads to higher contributions than with random re-matching. Page et al. (2005) regroup subjects after each third period according to their expressed preferences. This regrouping procedure increased contributions, both in a simple public goods setting and also in a public goods setting with punishment. In Cinyabuguma et al. (2005), subjects vote on irreversibly excluding others. Subjects who are not expelled contribute highly. Over time more and more subjects are banished which leads to a decrease in social efficiency. In a study by Charness and Lei-Yang (2008), subjects can decide whether they want to exit a group, to exclude other players by majority vote, or to merge with other groups (if 60% of the merging groups' members accept) before they play a public goods game. The greater the group size, the higher is the marginal social value, i.e., greater groups are more productive. Opportunities of regrouping together with the increase in the marginal social value raise contributions substantially. The increase in contributions, however, is smaller when the marginal social value is capped at a certain group size. Ahn et al. (2008) also investigate the group formation

⁴ Partner selection and its effects on behavior are also explored in market interactions (Kirchsteiger et al., 2005; and Brown et al., 2004), and networks (Riedl and Ule, 2003). Hauk and Nagel (2001) study the pure effect of unilateral and mutual choice of partners in finitely repeated prisoners' dilemma games.

under different rules: free entry and exit into a group, restricted (free) entry into and free (restricted) exit from a group. In the restricted entry and exit treatments, a majority vote decides on the entry and exit wish. Ahn et al. (2008) do not find significant differences between overall contributions. The different treatments, however, lead to groups of different sizes. Restricted entry leads to mid-sized and more effective groups while restricted exit and free entry/exit treatments lead to rather large and uncooperative groups. Ahn et al. (2009) also investigate the effect of the same rules in a congestible public good setting. For moderate group sizes (4-6 individuals), the restricted entry rule induces the highest contributions to the public good.

In all these studies subjects include or exclude interaction partners in static institutional settings. In contrast, in our voting with feet mechanism subjects do not actively choose their interaction partners, but opt for different institutions, that allow or do not allow for punishment. This at best permits an indirect choice of interaction partners since choosing an institution means to interact with “like-minded” individuals in the sense that all have chosen the same institution. Additionally, in contrast to our study, in the mentioned studies, productivities of groups increase with their size which is likely to make cooperation easier in larger groups.

Few studies focus on situations in which institution choice is the result of a voting process. Decker et al. (2003) let subjects vote for different punishment mechanisms. After observing others' contributions, subjects individually submit their intended punishment level for each of the other three group members. Whether the highest, the medium, or the lowest punishment level should be applied is auctioned off in a first-price auction. Subjects prefer the rule implementing the lowest of the proposed punishment levels which can be interpreted as unanimity vote on punishment, i.e., all agree that the respective player should be punished at least as much as the lowest level that is suggested. Ertan et al. (2009) allow subjects to interact in non-punishment and punishment institutions before they decide by majority vote whether punishment shall be available in a future interaction and if yes to whom it can be directed. Many groups opt to restrict punishment against high contributors but allow punishment of low contributors. These groups achieve higher levels of cooperation than groups with unrestricted or no punishment. Kroll et al. (2007) let subjects vote on a non-binding minimum contribution. This mechanism alone, however, turns out not to be effective. Only if contributions below the non-binding minimum contribution can be punished or the

minimum contribution becomes binding, contributions go up. In all three studies, players are not allowed to repeatedly choose to play the public goods game under different institutions so that self-selection into different institutions cannot occur. In Dal Bó et al. (2010) subjects play a series of prisoners' dilemma games and may choose an option that puts a fine on unilateral defection. It turns out that the punishment policy has a greater impact on cooperation (40% higher) if it is chosen by subjects' vote than when it is installed exogenously. Putterman et al. (2011) experimentally study a setting where subjects may vote for the adoption of a central punishment scheme consisting of two elements. Although many mechanisms are possible most groups vote for highly efficient mechanisms.

Two studies on the endogenous formation of institutions are most related to ours. Kosfeld et al. (2009) let players decide whether they are willing to participate in a sanctioning organization. All participants who declare their willingness to join the organization have to vote by unanimity rule whether the organization is actually implemented or not. If the organization is implemented, members are sanctioned for not contributing their full endowment. Outsiders of the organization are not sanctioned but nevertheless benefit from the contributions of all players. The data shows that a large majority of groups implement an organization by the final periods, despite the fact that it is costly. About 75% of these organizations even involve *all* players. A comparison with control treatments reveals that the opportunity to form organizations enhances group welfare by higher and stabilized contributions. Our approach is different in several aspects. First, we implement a setting in which punishment is decentralized. Therefore, our results do not rely on the creation of a global organization. Second, our public goods are local in the sense that players benefit from each others' contributions only within the same institution. Players joining different institutions play different public goods games.

Sutter et al. (2010) systematically investigate the effects of institution choice by voting between standard public goods, public goods with rewards, or public goods with punishment. Voting is costly, but not mandatory. Those who choose to vote repeat voting until they reach an unanimous decision. The vote determines the institution, under which *all* individuals subsequently have to interact for 10 periods, i.e., play includes also those players who decided not to vote. In different treatments the leverages, i.e., effectiveness of punishment and rewards are varied. For comparison, "exogenous treatments" are conducted, in which one of the institutions is exogenously imposed by the experimenter. Sutter et al. (2010) find that the

reward institution is chosen almost exclusively, particularly when rewards and punishment have a high leverage. The punishment institution is rarely chosen and only when the leverage is low. When it is selected, however, it is the most successful institution in eliciting high contributions. This is true in comparison with all other endogenously selected institutions, but also in comparison with all exogenously imposed institutions. The study by Sutter et al. (2010) analyzes choices between larger varieties of institutions, while we are more interested in the dynamic aspects of a voting with feet mechanism. Thus, we focus on two institutions, a standard public goods setting and one with sanctioning possibilities. Sutter et al. (2010) and Kosfeld et al. (2009) also differ in other respects from our approach. First, in their studies players cannot “escape” from each other, i.e., they are exogenously grouped together to play the public goods game. We allow players to choose institutions in every period which enables players to enter and exit. This leads to a dynamic evolution of the group composition. Thereby, we model an environment in which players have the freedom to escape, for example, if they are not satisfied with the institution or the treatment they receive from their respective partners. Secondly, we allow for a larger population which is three times as large as in the two other studies. Thus, in this respect we address a situation in which cooperation is particularly hard to achieve. Additionally, communities under each institution can grow endogenously and have a varying size and composition which is likely to make coordination on cooperative behavior even more difficult.

3. Community choice by voting with one’s feet

In our model, twelve players choose between the two communities before interacting in a voluntary contributions situation with others who have chosen the same community. The non-punishment community (N) resembles the standard voluntary contribution mechanism. In the sanctioning community (S), players may additionally engage in costly sanctioning of other players after having observed their contributions. We investigate two versions of sanctioning: punishment of others (P) or rewarding of others (R). For the description of the game, we will speak of sanctioning (S) in general, keeping in mind that it will be either in the form of punishment or in the form of rewarding. We consider a three-stage game consisting of a “voting with feet” stage (T0), a voluntary contribution stage (T1) and a sanctioning stage (T2).

3.1. “Voting with feet” stage (T0)

In T0 all twelve members of the population simultaneously choose one of the two

communities N and S. Once all choices are completed, each player is informed about the number of players n_θ , $\theta \in \{N, S\}$, who have chosen the same community. Notice that only the number of players and not their identities or histories of their play are revealed.

3.2. Contribution stage (T1)

In T1, a player i interacts only with those players who have chosen the same community. Each player is endowed with 20 monetary units (tokens) and may contribute an integer amount of g_i ($0 \leq g_i \leq 20$) to a joint project. Players decide simultaneously on their own contribution. The amount not contributed remains in the player's private account.⁵ The sum of all contributions $G_\theta = \sum_{j=1}^{n_\theta} g_j$ is multiplied by a_θ (with $1/n_\theta < a_\theta < 1$) and then consumed by each player, independent of the individual contribution g_i . The *marginal per capita return* (MPCR) a_θ – the return of each player from her own and others' contributions – depends on the community size n_θ . In order to give smaller communities the potential to be as productive as larger ones, we keep the *productivity* $R = n_\theta a_\theta$ constant for different group sizes by setting $R = n_\theta a_\theta = 1.6$. Thus, the return from the joint project for a player does not vary in n_θ if all members of a community symmetrically contribute a certain amount. As a consequence a_θ is a decreasing function in n_θ , as shown in Table 1.

Table 1: Marginal per capita return a_θ

Community size n_θ	2	3	4	5	6	7	8	9	10	11	12
Marginal per capita return a_θ	0.80	0.53	0.40	0.32	0.27	0.23	0.20	0.18	0.16	0.15	0.13

After all players have taken their contribution decisions, they are informed about the individual contributions of each member in their own community, again without revealing identities.

3.3. Sanctioning stage (T2)

In T2 each player receives 20 additional monetary units independent of her community affiliation and her contribution in T1. Providing players in N with the same additional

⁵ 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 T1 and T2.

endowment eliminates incentives to choose S just for receiving the extra tokens from T2. For the members of N, in this period the game ends here. The total monetary payoff of player i in N is

$$(1) \quad x_i^N = (20 - g_i + a_N G_N) + 20.$$

In S all players simultaneously decide whether or not to sanction other members of their community. All players are provided with the same sanctioning capacity, which is independent of their contributions g_i in T1. In total, each player may allocate up to 20 tokens.⁶ Player i can sanction community member j by assigning sanction tokens t_{ij} . Each token assigned by player i to player j incurs a cost of 1 token for player i . In case of punishment, one assigned token reduces the payoff of player j by 3 tokens in VF-P3R1 and in VF-P3. In case of rewarding an assigned token increases the payoff of player j by 1 token in VF-P3R1 and in VF-R1 and by 3 tokens in VF-R3. Let τ^i denote the amount of tokens that player i assigns and τ^{-i} denote the amount of tokens that player i receives from the other community members. The total monetary payoff of player i in S results in

$$(2) \quad x_i^S = \begin{cases} (20 - g_i + a_S G_S) + (20 - \tau^i - 3\tau^{-i}), & \text{for punishment with 1:3 in VF - P3R1 and in VF - P3} \\ (20 - g_i + a_S G_S) + (20 - \tau^i + \tau^{-i}), & \text{for reward with 1:1 in VF - P3R1 and in VF - R1} \\ (20 - g_i + a_S G_S) + (20 - \tau^i + 3\tau^{-i}), & \text{for reward with 1:3 in VF - R3} \end{cases}$$

The expressions in parentheses represent the stage payoffs of T1 and T2, respectively. After T2 is completed, all players are informed about all other players' contributions, their sanctioning tokens assigned, their sanctioning tokens received and their resulting total payoffs. We model an open information flow between both communities, i.e., at the end of each period, members of S are informed about the contributions and payoffs in N and members of N receive the same information about S as members of S do. The game described

⁶ We give subjects extra tokens for sanctions to separate the contribution choice from the sanctioning decision and to replicate the design of Güreker et al. (2006). The analysis of the actually assigned punishment and reward tokens in that study as well as in this paper provides no hint on an experimenter demand effect. It is worth noting that the threat of punishment is likely to be heavier in larger communities, as in total there are more punishment tokens available. In fact it is possible that a subject can be punished so heavily that she incurs a negative payoff in a period. To avoid bankruptcy, we provided subjects with a starting capital. Indeed, in the initial phase of the experiment, a number of subjects obtained negative period payoffs but overall no subject ran bankrupt.

so far is repeated for 30 periods involving the same twelve subjects.⁷

Joint payoffs are maximized when all community members fully cooperate (i.e., $g_i = 20$), no player punishes in S of VF-P3 and all players exert all reward tokens in S of VF-R3. Then each player's payoff is 52 tokens in N, in S of VF-P3 and in S of VF-R1 and 72 in S of VF-R3.

However, in the N-community contributing zero is the dominant strategy for money-maximizing players. Thus, in the unique Nash equilibrium, each player free-rides and earns 40 tokens. In S, a purely money-maximizing player will not sanction in T2 since it is costly to do so. Rational players foresee this and refrain from contributing in T1. Hence, independent of the community size, players do not contribute and do not sanction in the subgame-perfect equilibrium. In this case the total payoff is 40 tokens. Thus, in a world of money-maximizing rational actors, a player is indifferent between N and S because in equilibrium identical payoffs of 40 will be achieved, regardless whether the S institution entails a reward or a punishment mechanism. Since our interaction is finite, backward induction suggests that in our repeated setting, players will neither contribute nor sanction. Even if one assumes other-regarding preferences, like inequality-averse players in the spirit of Fehr and Schmidt (1999), one comes to similar results (see Appendix B).

4. Experimental Design and Procedure

Subjects were recruited for voluntary participation via the online recruitment system ORSEE (Greiner, 2004) and were randomly allocated to treatments. None of them had participated in a similar experiment before. On arrival, subjects were informed about the experimental procedure as well as the number of periods.⁸ The experiment was programmed with *z-Tree* (Fischbacher, 2007) and conducted in the computerized laboratory *eLab* at the University of Erfurt. Random reshuffling of the presentation order on the computer screens ensured that the identity of the players could not be traced over periods. In total 456 subjects participated, i.e., we collected 38 independent observations (8 in VF-R1, 8 in VF-R3, 8 in VF-P3, 8 in GX-P3, and 6 in FX-P3) with 12 subjects each. An experimental session lasted for about 2 to 2.5

⁷ In our experiment, we consider a special case of a partner design in which not all members of the group necessarily interact in each period. 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).

⁸ A translation of the instruction sheet is given in Appendix A. Original instructions were written in German. They are available upon request from the authors.

hours. Tokens gained were converted with an a priori known conversion rate into money. Subjects earned between 15 and 25 Euro.

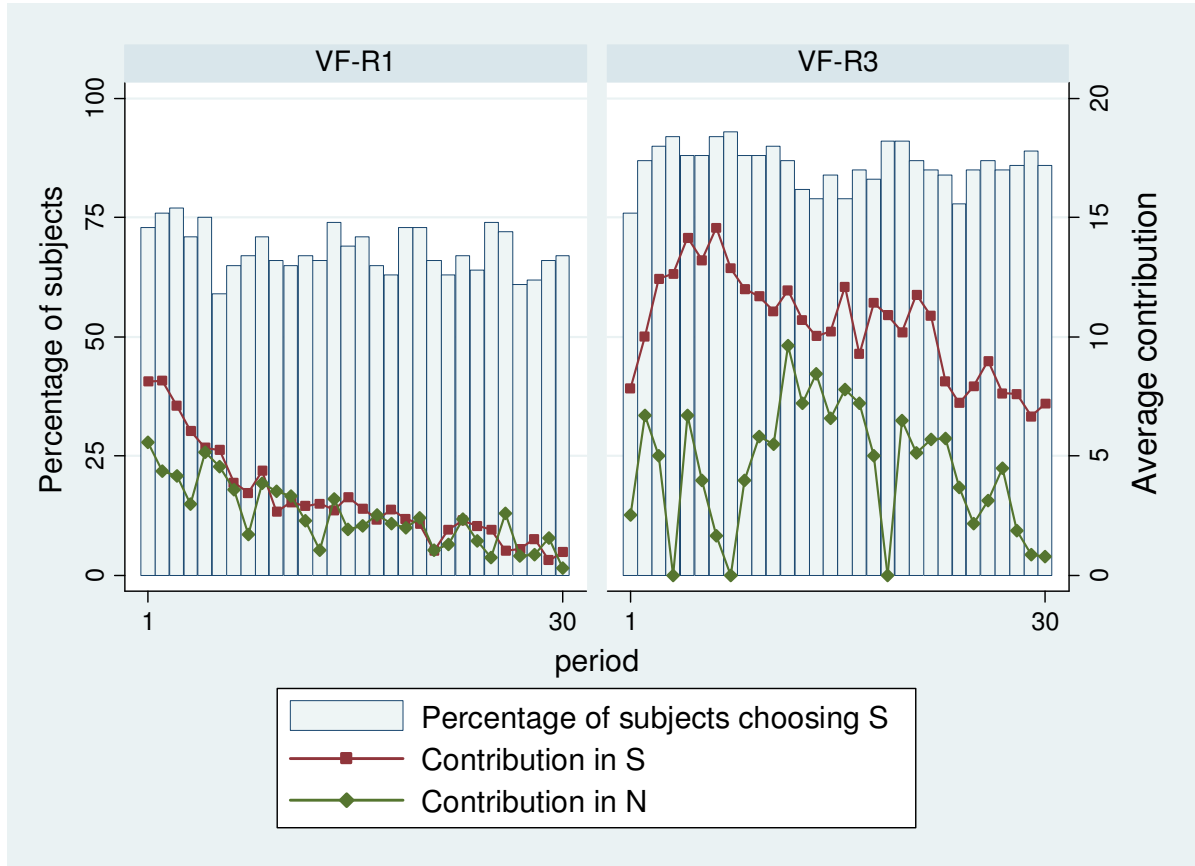
5. Results

5.1. Reward experiments VF-R1 and VF-R3

Figure 1 and Figure 2 show that the institution choice in the reward experiments is distinctively different from VF-P3R1 in two important aspects. First, already at the start, the vast majority of subjects (about 70% in VF-R1 and about 80% in VF-R3) opt for the reward institution. Second, the percentage of subjects choosing the reward institution is roughly the same in each period. The high acceptance, however, does not translate into very high contributions. Statistical tests show that the contributions in S of VF-P3R1 are significantly higher than the contributions in S of VF-R1 ($p < 0.000$) and in S of VF-R3 ($p < 0.001$).⁹ Average contributions in S of VF-R1 decrease right from the beginning while in S of VF-R3 contributions initially increase. From period 8, however, the trend reverses and we observe a steady decrease in contributions in S also in the high leverage institution VF-R3. Thus, the data from the two reward treatments clearly show that endogenous choice between a non-sanction and a reward institution does not stabilize contributions, even if rewarding is extra profitable by being efficiency enhancing.

⁹ All non-parametric statistical tests reported in this paper are two-tailed and take communities as units of observations. Comparisons within a treatment are tested with the Wilcoxon matched pairs tests. Comparisons across treatments are performed with the Whitney Mann U-tests.

Figure 1: Comparison of VF-R1 and VF-R3

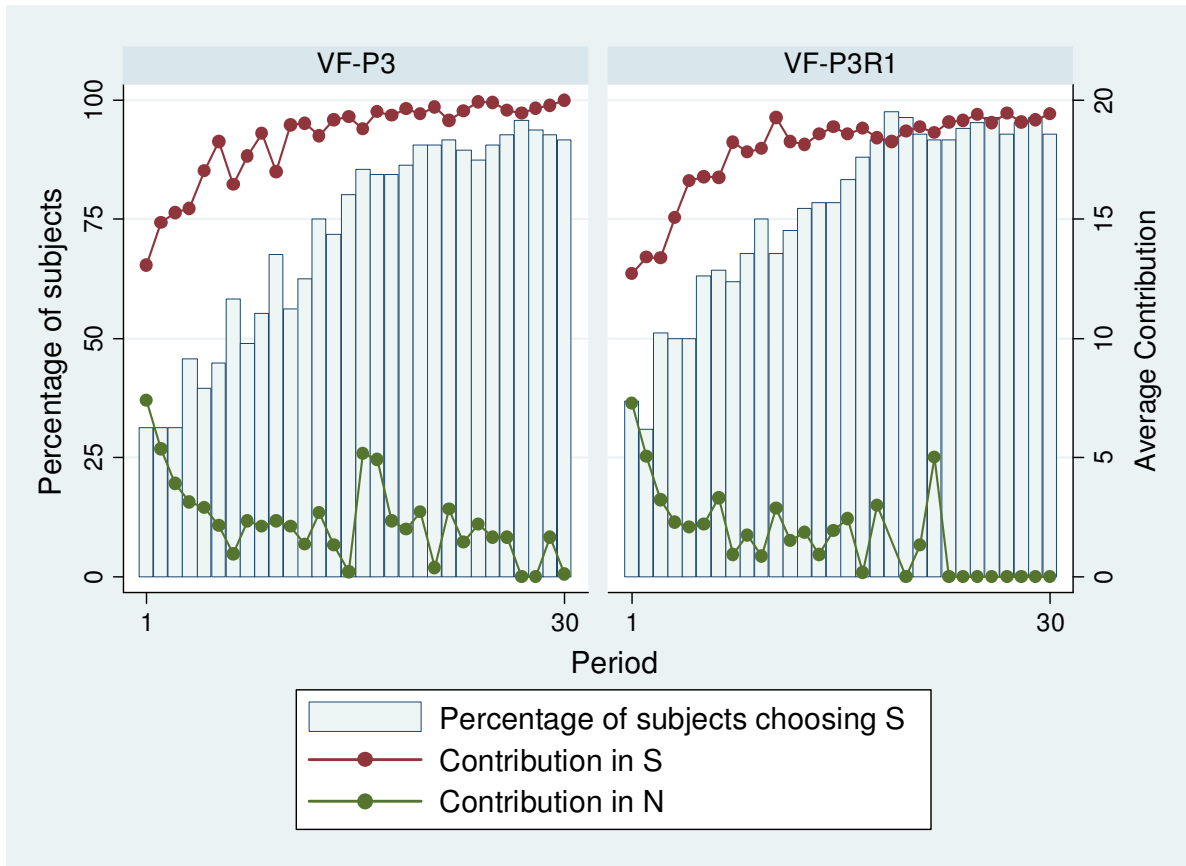


5.2. Voting with feet experiment VF-P3

The experimental sessions on voting with feet experiment between a punishment and no-sanctioning institution qualitatively yields the same results as in VF-P3R1 where the sanctioning institution allows for both, punishment *and* reward (see Figure 2). Neither initially nor over time there are significant differences between VF-P3R1 and VF-P3 regarding the main comparisons: community choice, contribution and punishment behavior.¹⁰

¹⁰ There are no significance differences in contributions between VF-P3R1 and VF-P3 (initial in N: $p = 1.000$; initial in S: $p = 0.608$; overall in N: $p = 0.315$; overall in S: $p = 1.000$). The number of subjects does not differ significantly between VF-P3R1 and VF-P3 (initial number of subjects in N: $p = 0.315$; initial number of subjects in S: $p = 1.000$; overall number of subjects in N: $p = 0.315$; overall number of subjects in S: $p = 0.132$). Finally, the received punishment in S is not significantly different between VF-P3R1 and VF-P3 (initial $p = 0.103$; overall $p = 0.103$).

Figure 2: Comparison of VF-P3 and VF-P3R1



What makes the voting with feet mechanism between the non-sanctioning and a punishment institution so successful? In the light of our results, two explanations are at hand. The first is that *initial* self-selection into communities is key for success. It allows the subjects to join groups of “like-minded” people who might be convinced by the potential of a punishment institution to establish a cooperative “culture”. The other non-exclusive explanation is the growth process of the punishment community, i.e., the mere fact that the community starts off with a small group of subjects and then grows to its maximum size. To investigate and to disentangle these two possible explanations we ran two control experiments, which will be described and discussed in the two following sections.

5.3. The effect of self-selection

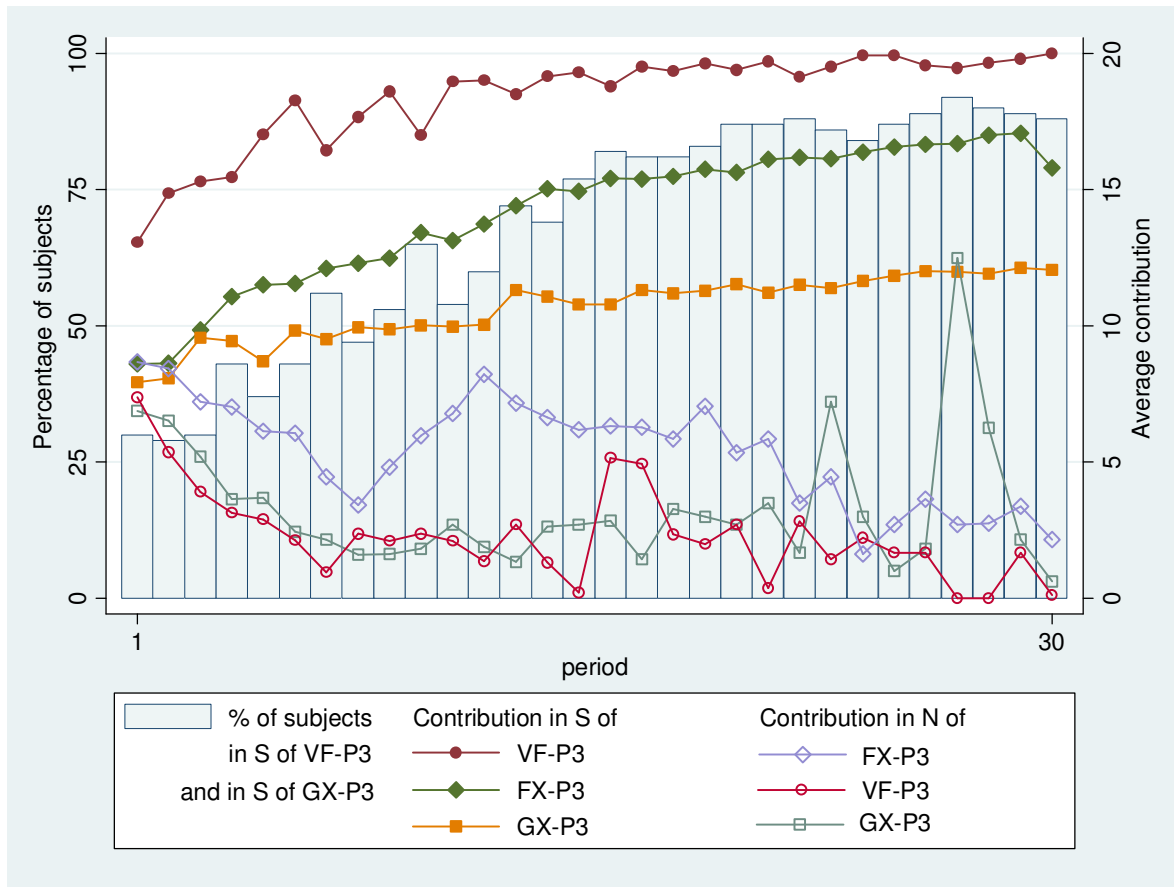
We report a control experiment designed with identical growth paths as in VF-P3, but without the possibility of self-selection. Instead, we exogenously allocate subjects to the communities N and S. We refer to this experiment as the exogenous growth experiment (abbreviated GX-

P3). In GX-P3 there is no community choice stage. Instead, all community choice vectors of subjects who participated in one session of the VF-P3 experiment are randomly assigned to participants in one session of the GX-P3. Technically, each session in GX-P3 exactly re-runs the community choices of the “mirror”-session in VF-P3. We conducted 8 additional sessions with 12 new subjects each. Subjects were told that community affiliations may vary from period to period and will be announced privately at the beginning of a period. As in VF-P3, subjects are informed on the community size, but not on the identities of their members.

The initial periods

First period contributions in S of VF-P3 (13.2 tokens) are significantly higher than first period contributions in S of GX-P3 (8.7 tokens, $p = 0.003$, cf. Figure 3). In contrast to VF-P3, in GX-P3, the first period contributions in S and N are quite similar. In both N and S of GX-P3, only 16.7% contribute high while 43.3% contribute low in S and 59.1% in N. Compared to VF-P3, the initial punishment behavior is also systematically different than in GX-P3. In the first period, we observe only one subject who contributes high and punishes a less-contributor. This is significantly less than in VF-P3 where 14 subjects are high contributors who punish less-contributors ($p = 0.004$). Additionally, in the first period, allocated punishment to less-contributors (per token the other contributed less) is significantly higher in VF-P3 than in GX-P3 ($p = 0.051$). Hence, the disciplining subjects in S of VF-P3 are not only more numerous than in S of GX-P3, but they also show less mercy against low contributors.

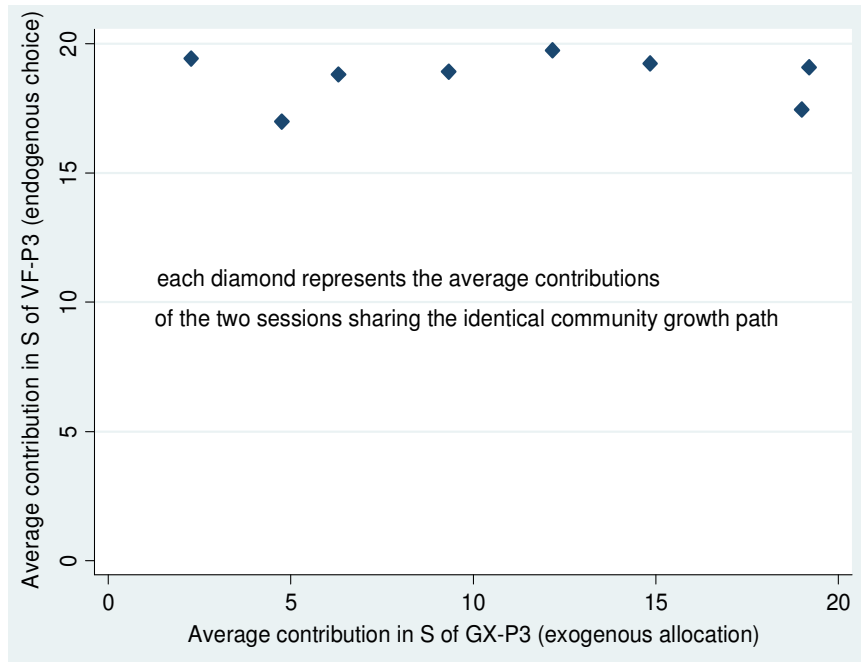
Figure 3: Community choice over periods and contributions in S and in N



Development over time

Figure 4 relates the average contribution of each S-community in VF-P3 to the average contribution of its “mirror” community in GX-P3 with identical (but exogenously imposed) growth-path. While the variability in contributions is low in S of VF-P3 (variance: 0.83) there is a huge difference in the variability of average contributions in S of GX-P3 (variance: 35.9). Overall, subjects in S of VF-P3 contribute significantly more than subjects in S of GX-P3 (18.7 tokens and 11.0 tokens, resp., $p = 0.016$). There is no significant difference between the overall contributions in N of VF-P3 (3.6 tokens) and in N of GX-P3 (3.9 tokens) ($p = 0.619$).

Figure 4: Average contributions in S of VF-P3 and in S of GX-P3



How do subjects in GX-P3 adapt their behavior after they have been moved from N to S? In GX-P3, 70.0% of the subjects increase their contribution after switching from N to S; in VF-P3 this is the case for 76.3% ($p = 0.372$). Thus, subjects seem to be well aware of the fact that behavior in S should be different than in N. The incoming subjects in S of GX-P3, however, increase their contributions less than subjects do in the same situation in VF-P3 (5.6 tokens and 10.3 tokens, resp., $p = 0.028$). Only 6.7% of the subjects in S of VF-P3 who are punished in period $t-1$ but choose to remain in S in period t , decrease their contributions. These are significantly fewer than in GX-P3 (15.6%) where the decision to remain in S is predetermined ($p = 0.062$). If the subjects, who have been punished in period $t-1$ and stay in S, increase their contributions in period t , then this increase is significantly lower in S of GX-P3 than in S of VF-P3 (3.6 tokens and 5.3 tokens, resp., $p = 0.065$). Punishment behavior also differs between both treatments. Punishment allocated to a less-contributor (per token the other contributed less) is higher in VF-P3 than in GX-P3 ($p = 0.012$). Also, on average, a punisher in S of GX-P3 contributes significantly less than a punisher in S of VF-P3 (11.1 and 17.7 tokens, resp., $p = 0.005$). As a consequence of contribution and punishment behavior, in the second half of the experiment, payoffs in S of GX-P3 are weakly significantly lower than payoffs in S of VF-P3 ($p = 0.099$).

5.4. The effect of community growth

Another characteristic of VF-P3 is the steady growth of the S communities, from a minority to

the entire population. To study the pure effect of this community growth on cooperation we report a control experiment with new subjects and with fixed sized communities (abbreviated FX-P3). In this control experiment there again is no community choice stage. Instead, in each of the 6 sessions two equally sized communities (S and N) are established consisting of 6 members each. The community composition is stable throughout the session. Comparing FX-P3 to GX-P3 provides us with insights about the impact of the community growth on cooperation.

First period contributions in S of GX-P3 and S of FX-P3 are not significantly different ($p = 0.564$). The same is true for N of GX-P3 and N of FX-P3 ($p = 0.169$). Also overall contributions between S and N-communities of GX-P3 and of FX-P3 are not significantly different ($p = 0.414$ and $p = 0.513$, respectively). Within each experiment, overall contributions in S are significantly higher than in N (VF-P3: $p = 0.008$, GX-P3: $p = 0.008$, and FX-P3: $p = 0.031$, respectively).

With respect to punishment of less-contributors there is no significant difference between S of GX-P3 and S of FX-P3. Punishment allocated to a less-contributor (per token the other contributed less) is not significantly different between GX-P3 and FX-P3 – neither in the first period ($p = 0.394$) nor overall ($p = 0.108$). Overall, there is also no difference in contribution behavior of punishers between both treatments ($p = 0.491$). In the first half, there is no difference between payoffs in S of GX-P3 and payoffs in S of FX-P3 ($p = 1.000$). In the second half, however, payoffs in S of FX-P3 are weakly significantly higher than payoffs in S of GX-P3 ($p = 0.087$). These results suggest that the mere slow growth effect of an S community that starts small and increases steadily is not substantial.¹¹

6. Summary and conclusion

In this study, we investigate potential reasons for the cooperation and efficiency enhancing effect of the voting with feet (VF) mechanism observed in Gürer et al. (2006). We show that the effect is not driven by the interaction of punishment and reward in the sanctioning institution. In fact, punishment alone has the same contribution enhancing effect, while pure reward possibilities do not sustain cooperation in a VF setting, even if rewards create additional efficiency gains.

¹¹ It might well be that a possible advantage of growing slowly in GX-P3 might be canceled out by a possible advantage of having a fixed group composition in FX-P3 (which arguably makes it easier to sustain cooperation). However, we do not have any conclusive indications pointing into this direction.

To further identify the determinants of the success of VF we separate between the self-selection and the slow growth explanation in a VF setting with a choice between a non-sanction institution and a punishment institution. In the GX-P3 treatment, we show that communities with exogenous subject allocation and identical growth paths as in VF-P3 perform significantly less successful. With the help of a second control experiment (FX-P3) we find that growing communities per se are not more successful than fixed-size ones, if subject allocation is exogenous. This suggests that *initial* endogenous self-selection of subjects is an important key for the establishment and efficient maintenance of cooperation. In the beginning the punishment community attracts subjects who contribute high and harshly punish defectors. Although entry and exit is not restricted, high cooperation levels are established and sustained. Strikingly, this is even true when the entire subject population ultimately joins S and despite the fact that MPCRs in S become very small while MPCRs become large in N (due to the small group size).

Our findings highlight a so far undervalued feature of the voting with feet mechanism: In our VF setting consumers do not choose between communities with different public goods but they choose between communities with different institutional rules that are intended to govern the provision of the public goods.¹² Thus, in addition to the efficiency improvement from the “consumer-voter [...] picking that community which best satisfies his preference pattern for public goods” as suggested by Tiebout (1956), the VF mechanism improves efficiency by facilitating the right initial match between consumers and different institutional rules.

In our voting with feet setting the punishment institution seems to serve at least two purposes. First, it provides effective tools to discipline low contributors. Second, the punishment

¹² A notable exception is Romer’s concept of “charter cities” that recently has been chosen as one of the 10 breakthrough ideas for 2010 by the Harvard Business Review in collaboration with the World Economic Forum (Romer, 2010a, 2010b). In a nutshell the idea behind the concept is that economic growth is likely to be supported by giving people the possibility to *freely* choose (to work/live) among (co-existing) jurisdictional territories governed by different institutional rules. The rules of charter cities may not only differ in degree (like different tax rates) but even in quality like having completely different laws for commerce and civil law. A key point of charter cities (different from colonialism) is that there is no coercion to move. Everyone is free to choose under which jurisdiction he or she wants to live and hence (implicitly) accepts and gets involved under the rules that govern the particular jurisdiction. As stated by Romer on his website “[t]he process of movement between can be more effective than the process of change from within” (<http://www.chartercities.org/>). An example of a charter city is Hong Kong which has different institutional rules than mainland China. In Hong Kong Chinese people – different than their fellow citizens in the mainland – could live and work under capitalistic market structure which resulted in an enormous economic growth in the second half of the 20th century. The success of the initially small city convinced more and more citizens to leave mainland China and to move to Hong Kong. See also Frey and Eichenberger (1999) for a related concept of functional, overlapping, and competing jurisdictions.

institution initially attracts subjects who behave cooperatively and punish. Indeed the initial members contribute higher and punish more severely than subjects who are exogenously allocated to a punishment institution (c.f. the comparison between VF-P3 and GX-P3). Although our study does not allow for clearly disentangling different reasons of this observation it seems to be plausible that in the beginning the punishment institution serves as a coordination device for subjects with an inclination to contribute high and punish. These subjects might foresee and appreciate the potential of the punishment institution to enforce cooperation. Cooperation and punishment is likely to be reinforced by the voting with feet mechanism because subjects in the punishment institution might expect to interact with others who are like-minded. It seems that the self-selected founding members of the punishment institution are able to create and establish an initial culture which is decisive for the behavior of subsequently incoming members.¹³

Our findings also raise important questions for future research that are beyond the scope of this paper. For feasibility reasons we concentrate on two specific institutions. It would be interesting to design various alternative institutions and to see how successful they are in a voting with feet mechanism. Introducing a larger variety of institutions would certainly come closer to Tiebout's vision. How would a larger variety of institutions affect cooperation success of the voting with feet mechanism? Dealing with these questions would help to understand the potentials and drawbacks of voting with feet mechanisms and to design voting with feet institutions that mitigate the downsides of social dilemmas.

¹³ Founding entrepreneurs, for example, have a very strong influence on the culture of the organization they create (Schein 2004).

References

- Abbink, K., Irlenbusch, B., Renner, E., 2000. The Moonlighting Game – An Experimental Study of Reciprocity and Retribution. *Journal of Economic Behavior and Organization* 42, 265-277.
- Ahn, T.-K., Isaac, M., Salmon, T.C., 2008. Endogenous Group Formation. *Journal of Public Economic Theory* 10(2), 171-194.
- Ahn, T.-K., Isaac, M., Salmon, T.C., 2009. Coming and going: Experiments on endogenous group sizes for excludable public goods. *Journal of Public Economics* Volume 93, 336-351.
- Andreoni, J., Harbaugh, W., Vesterlund, L., 2003. The Carrot or the Stick: Rewards, Punishments, and Cooperation. *American Economic Review* 93(3), 893-902.
- Balliet, D., Mulder, L. B., van Lange, P. A. M. 2011. Reward, Punishment, and Cooperation: A Meta-Analysis. *Psychological Bulletin*, 137(4), 594–615.
- Blanco, M., Engelmann, D., Normann, H.-T., 2010. A Within-Subject Analysis of Other-Regarding Preferences. Working Paper, University of Düsseldorf.
- Bowles, S, 2004. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton University Press.
- Brekke, A., Hauge, K.E., Lind, J.T., Nyborg, K. 2011. Playing with the good guys. A public good game with endogenous group formation. *Journal of Public Economics* 95, 1111-1118.
- Brown, M., Falk, A., Fehr, E., 2004. Relational Contracts and the Nature of Market Interactions. *Econometrica* 72, 747-780.
- Carpenter, J., 2007. Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods. *Games and Economic Behavior* 60(1), 31-52.
- Charness, G., Yang, C.-L., 2008. Endogenous Group Formation and Public Goods Provision: Exclusion, Exit, Mergers, and Redemption. Working Paper, UCSB.
- Cinyabuguma, M., Page, T., Putterman, L., 2005. Cooperation under the Threat of Expulsion in a Public Goods Experiment. *Journal of Public Economics* 89, 1421-1435.
- Coricelli, G., Fehr, D., Fellner, G., 2004. Partner Selection in Public Goods Experiments. *Journal of Conflict Resolution* 48(3), 356-378.

- Croson, R.T.A., 1996. Partners and strangers revisited. *Economics Letters* 53, 25-32.
- Croson, R.T.A., 1998. Theories of Altruism and Reciprocity: Evidence from Linear Public Goods Games. Working Paper 98-11-04, The Wharton School of the University of Pennsylvania.
- Dal Bó, P., Foster, A., Putterman, L., 2010. Institutions and Behavior: Experimental Evidence on the Effects of Democracy. *American Economic Review* 100, 2205-2229.
- Davis, D.D., Holt, C.A., 1993. *Experimental Economics*. Princeton University Press.
- Dawes, R.M., 1980. Social Dilemmas. *Annual Review of Psychology* 5, 163-193.
- Decker, T., Stiehler, A., Strobel, M., 2003. A Comparison of Punishment Rules in Repeated Public Good Games. *Journal of Conflict Resolution* 47(6), 751-772.
- Ehrhart, K.-M., Keser, C., 1999. Mobility and Cooperation: On the Run. Working paper 99s-24. CIRANO, Montreal.
- Ertan, A., Page, T., Putterman, L., 2009. Who to punish? Individual Decisions and Majority Rule in Mitigating the Free Rider Problem. *European Economic Review* 53, 495-511.
- Falk, A., Gächter, S., Fischbacher, U., in press. Living in Two Neighborhoods - Social Interaction Effects in the Lab. *Economic Inquiry*.
- Fehr, E., Schmidt, K.-M., 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114(3), 817-868.
- Fehr, E., Gächter, S., 2000. Cooperation and Punishment in Public Goods Experiments. *American Economic Review* 90(4), 980-994.
- Fehr, E., Gächter, S., 2002. Altruistic Punishment in Humans. *Nature* 415, 137-140.
- Fischbacher, U., Gächter, S., 2010. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments. *American Economic Review* 100(1), 541-556.
- Fischbacher, U., 2007. z-Tree: Zurich Toolbox for Ready-made Economic experiments. *Experimental Economics* 10(2), 171-178.
- Frey, B. S., Eichenberger, R., 1999. *The new democratic federalism for Europe: functional, overlapping, and competing jurisdictions*. Elgar, Cheltenham, UK.
- Gächter, S., Thöni, C., 2005. Social Learning and Voluntary Cooperation among Like-Minded People. *Journal of the European Economic Association* 3, 303-314.

- Gächter, S., Renner, E., Sefton, M., 2008. The Long-Run Benefits of Punishment. *Science* 322, 1510.
- Greiner, B., 2004. An Online Recruiting System for Economic Experiments. In: Kurt Kremer, Volker Macho (eds.). *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*, Göttingen: Ges. Für Wiss. Datenverarbeitung, 79-93.
- Gürerk, Ö., Irlenbusch, B, Rockenbach, B., 2006. The Competitive Advantage of Sanctioning Institutions. *Science* 312, 108-111.
- Hardin, G., 1968. The Tragedy of the Commons. *Science* 162, 1243-1248.
- Hauk, E., Nagel, R., 2001. Choice of Partners in Multiple Two-Person Prisoner's Dilemma Games. *Journal of Conflict Resolution* 45(6), 770-793.
- Isaac, M., Walker, J., 1988. Group Size Hypotheses of Public Goods Provision: An Experimental Examination. *Quarterly Journal of Economics* 103, 179-199.
- Keser, C., van Winden, F., 2000. Conditional Cooperation and Voluntary Contributions to Public Goods. *Scandinavian Journal of Economics* 102(1), 23-39.
- Kirchsteiger, G., Niederle, M., Potters, J., 2005. Endogenizing Market Institutions: An Experimental Approach. *European Economic Review* 49(7), 1827-1852.
- Kosfeld, M., Okada, A., Riedl, A.M., 2009. Institution Formation in Public Goods Games. *American Economic Review* 99(4), 1335-1355.
- Kroll, S., Cherry, T.L., Shogren, J.F., 2007. Voting, Punishment, and Public Goods. *Economic Inquiry* 45(3), 557-570.
- Ledyard, J., 1995. Public Goods: A Survey of Experimental Research. J. Kagel and A. Roth, *Handbook of Experimental Economics*, Princeton University Press, 111-194.
- Milinski, M., Rockenbach, B., On the interaction of the stick and the carrot in social dilemmas. *Journal of Theoretical Biology*, in press, [doi:10.1016/j.jtbi.2011.03.014](https://doi.org/10.1016/j.jtbi.2011.03.014).
- Ostrom, E., 1998. A Behavioral Approach to the Rational Choice Theory of Collective Action. *American Political Science Review* 92(1), 1-23.
- Ostrom, E., 1999. Coping With the Tragedy of the Commons. *Annual Review of Political Science* 2, 493-535.
- Ostrom, E., Walker, J., Gardner, R., 1992. Covenants with and without a Sword: Self-Governance is Possible. *American Political Science Review* 86, 404-417.

- Page, T., Putterman, L., Unel, B., 2005. Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency. *Economic Journal* 115, 1032–1053.
- Putterman, L., Tyran, J.-R., Kamei, K., 2011. Public goods and voting on formal sanction schemes. *Journal of Public Economics* 95, 1213-1222.
- Riedl, A.M., Ule, A., 2003. Exclusion and Cooperation in Social Network Experiments. Working Paper, University of Amsterdam.
- Romer, P.M., 2010a. Creating More Hong Kongs. *Harvard Business Review* 88(1/2), 55-56.
- Romer, P.M., 2010b. What Parts of Globalization Matter for Catch-Up Growth? *American Economic Review* 100(2), 94-98.
- Schein, E.H., 2004. *Organizational Culture and Leadership*. Jossey-Bass, San Francisco.
- Sefton, M., Shupp, R., Walker, J., 2007. The Effect of Rewards and Sanctions in the Provision of Public Goods. *Economic Inquiry* 45, 671-690.
- Sutter, M., Haigner, S., Kocher, M. (2010). Choosing the stick or the carrot? – Endogenous institutional choice in social dilemma situations. *Review of Economic Studies* 77, 1540-1566.
- Tiebout, C.M., 1956. A Pure Theory of Local Expenditures. *Journal of Political Economy* 64(5), 416-424.
- Time Magazine, 1955. Monday, 21 November. From the article: GENEVA: Cold Finalities. <http://www.time.com/time/magazine/article/0,9171,866620,00.html>
- Weber, R.A., 2006. Managing growth to achieve efficient coordination in large groups. *American Economic Review* 96(1), 114-126.
- Yamagishi, T., 1986. The Provision of a Sanctioning System as a Public Good. *Journal of Personality and Social Psychology* 51(1), 110-116.

Appendix A: Instructions for the experiment (Treatment VF-P3)¹⁴

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** = $1.6 \times \text{sum of the contributions of all group members} / \text{number of group members}$

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

20 – your contribution to the project

+ $1.6 \times \text{sum of the contributions of all group members} / \text{number of group members}$

¹⁴ We provide the instructions for the VF-P3 treatment. The instructions for the other treatments are written analogously. To save space, we do not include them here. They are available from the authors upon request.

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** 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

$$\begin{aligned} & \text{Your payoff from Stage 1} = 20 - \text{your contribution to the project} + 1.6 \times \frac{\text{sum of the contributions of all group members}}{\text{number of group members}} \\ + & \text{Your payoff from Stage 2} = 20 - \text{sum of the tokens that you have assigned to other group members} \\ & \quad - 3 \times (\text{the number of tokens you have received from other group members}) \end{aligned}$$

$$= \text{Your round payoff}$$

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.

We wish you success!

Appendix B: Theoretical Predictions with Preferences of Inequality-Aversion (not intended for journal publication but as supplementary material)

General Overview

Let us assume that (at least some) players are not exclusively motivated by their own monetary payoffs, but have other-regarding preferences relative to the members of their current community¹⁵; as modeled, for example, in Fehr and Schmidt (1999). They suggest a utility function where 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 . They assume $0 \leq \beta_i < 1$ and $\alpha_i \geq \beta_i$. If $x = (x_1, \dots, x_{n_\theta})$ denotes the vector of the individual monetary payoffs of the n_θ community members, player i 's utility is described by

$$(A1) \quad U_i(x) = x_i - \alpha_i \frac{1}{n_\theta - 1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n_\theta - 1} \sum_{j \neq i} \max\{x_i - x_j, 0\}.$$

Fehr and Schmidt (1999) apply their model to a public goods setting with voluntary contributions and show that equilibria may exist in which *conditional cooperators* contribute strictly positive amounts to the public good.¹⁶ In a first step, we consider the communities separately and examine the conditions for equilibria with positive levels of cooperation. The analysis of the static consideration of a community is an extension of Fehr and Schmidt's examination, taking the varying community sizes into account. In a second step, we extend this analysis by discussing the dynamics inherent in the endogenous choice process.

Analysis of the non-sanctioning community (N)

Adapting Fehr and Schmidt's original analysis to our model, we can show that players who are not sufficiently averse to advantageous inequality (i.e., players with parameters $\beta_i < 1 - R/n_N$) never contribute to the public good when sanctions are absent, as in our N

¹⁵ For simplicity we abstract from possible inequality aversion regarding members of the other community reflecting the fact that reference groups are often constituted by people with whom one directly interacts in a certain time span

¹⁶ A "conditional cooperator" is a player who reduces his disutility from advantageous inequality by contributing himself if other players also contribute. A necessary condition for tolerating some free-riders in a community without punishment is that the suffering from disadvantageous payoff inequality is not too high cf. Fehr and Schmidt (1999), Proposition 4 (c), p. 839.

community. In addition, when the productivity parameter $R < 2$, as it is the case in our experimental parameterization as well as in the vast majority of similar experiments, we can show that in the N community, equilibria with positive contributions exist if and only if *all* players i are sufficiently averse to advantageous inequality, i.e., $\beta_i \geq 1 - R/n_N$ for $i = 1, \dots, n_N$. In all these equilibria, all players contribute the same amount. Thus, when $R < 2$, we can be sure that there is no equilibrium with positive cooperation levels in N, if there is at least one single player who is not sufficiently averse to advantageous inequality. The proof of these statements is given below (Proposition 1).

In our experimental parameterization the lower bound for β_i necessary to enable cooperation in equilibrium varies between 0.20 (for $n_N = 2$) and 0.87 (for $n_N = 12$) and thus becomes the more demanding, the larger the community is. Fehr and Schmidt propose an average $\beta = 0.315$ and according to empirical estimations of the inequality aversion parameters (e.g., Fehr and Schmidt 1999, p. 844; and Blanco et al., 2010), it is almost impossible to observe β -values that are sufficiently high to allow cooperation in larger communities. Thus, equilibria with positive contributions are highly unlikely in N.

Analysis of the punishment community (P)

In their analysis of institutions with punishment possibilities, Fehr and Schmidt (1999) assume, for the sake of simplicity, that all players are of one of two types: *enforcers* who conditionally cooperate and are ready to punish deviators or *payoff maximizers* with inequity parameters $\alpha = \beta = 0$. The adaptation of Fehr and Schmidt's model to the punishment community shows that equilibria with positive contributions exist if some players suffer sufficiently from disadvantageous inequality and credibly threaten to punish free-riders. In these equilibria, *all* players contribute an identical amount.

We extend the analysis of Fehr and Schmidt and show that under reasonable assumptions equilibria with positive contributions can only exist if the number of payoff maximizers in the punishment community is low. To be precise, the number of payoff maximizers has to be strictly lower than the multiplicative inverse of the marginal punishment costs $1/c$. In our setting with $c = 1/3$ this implies that the punishment community can “afford” at most two payoff maximizers. If more than two payoff maximizers join the punishment community, no equilibrium with positive contribution levels exists. Interestingly, this is true independent of

the community size n_s . This means that no matter how many members join the punishment community, more than two payoff maximizers destroy the possibility of an equilibrium with positive contributions.

If there are no enforcers who threaten to punish the non-contributors, the situation is “equivalent” to the situation in N described above: equilibria with positive contributions only exist if all players are conditional cooperators with $\beta_i \geq 1 - R/n_p$, regardless how large the community size n_p is. A detailed analysis of above statements with proofs is given in Proposition 2 and Corollary 1.

Analysis of the reward community (R)

While both the analyses of N and the punishment community are adaptations of Fehr and Schmidt’s considerations, our analysis of the reward community has no master copy in Fehr and Schmidt.

We find that a (separating) equilibrium with different contributions for free-riders and conditional cooperators does not exist (Lemma 1 and Proposition 3). 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 (Lemma 2 and Proposition 4). Cooperation equilibria may also exist if all players are conditional cooperators.

If we take the parameter values suggested by Fehr and Schmidt (1999) or estimated by Blanco et al. (2010) the chances for observing cooperation in larger communities is extremely low (as in case of the punishment community).

Analysis of the dynamics

What does the Fehr-Schmidt model suggest for the dynamic case of voting-by-feet choice? Conditional cooperators who sufficiently dislike disadvantage inequality are likely to choose the punishment community because this gives them the possibility to (actively) establish the cooperation equilibrium by threatening to punish free-riders. In the reward community, conditional cooperators might be able to sustain an equilibrium in which all players contribute by rewarding contributors. Hence the sanctioning possibilities may well serve as a coordination device for conditional cooperators to gather in the sanctioning community S. If conditional cooperators vote with their feet for S, cooperation payoffs would likely be higher

in S than in N. This could attract other players to join S. However, with a growing community size, sanctioning another player becomes less attractive, since although it would equalize payoffs towards the sanctioned player, it would in fact increase the inequality towards the (many) players who contributed, but do not sanction. For all communities, the chances of the existence of a cooperation-equilibrium tend to diminish the larger the community size is or becomes over time. Thus, even when one assumes that players are inequity averse it is highly unlikely that cooperation emerges in our population of 12 players.

Propositions

Adaption of the Fehr and Schmidt (1999) results to the N community

Fehr and Schmidt (1999) show that in a public goods setting with voluntary contributions and no sanctioning possibility contributing zero is a dominant strategy for players with $a_N + \beta_i < 1$.¹⁷ Since in our model the MPCR varies with the community size while the productivity of the public good is constant, i.e., $R = n_N a_N$, we replace a_N by R/n_N . Thus, for players with $R/n_N + \beta_i < 1$ it is a dominant strategy to contribute nothing. In the following, we show that there is an equilibrium in N in which all players contribute zero and that equilibria with positive contributions exist only if *all* players are “conditional cooperators”, i.e., their preferences satisfy $R/n_N + \beta_i \geq 1$.

Proposition 1. Assume that $R < 2$.

- I. If for at least one player $R/n_N + \beta_i < 1$ is satisfied, the unique equilibrium in N prescribes free-riding of all players.
- II. Equilibria with strictly positive contributions exist if and only if *all* players i satisfy $R/n_N + \beta_i \geq 1$. In all these equilibria, all players contribute the same amount.

Proof.

- I. In analogy to Fehr and Schmidt (1999) assume that there are n_N players in N with contributions $g_1 \leq g_2 \leq \dots \leq g_{n_N}$ and $k > 0$ of these n_N players have $R/n_N + \beta_i < 1$. As mentioned above, for these k players it is a dominant strategy to contribute zero:

¹⁷ Cf. the proof to the Proposition 4, part (a) in the Appendix of Fehr and Schmidt (1999), p. 860.

$g_1 = \dots = g_k = 0$. Suppose there exists a player $l > k$ who contributes the smallest positive amount $0 = g_{l-1} < g_l \leq g_{l+1} \leq \dots \leq g_{n_N}$. Player l 's utility is given by

$$(A2) \quad U_l(g_l) = y - g_l + \frac{R}{n_N} g_l + \frac{R}{n_N} \sum_{j \neq l+1}^{n_N} g_j - \frac{\beta_l}{n_N - 1} \sum_{j \neq l+1}^{n_N} (g_j - g_l) - \frac{\alpha_l}{n_N - 1} \sum_{j \neq l}^{l-1} g_l.$$

Rearranging the terms we obtain:

$$(A3) \quad U_l(g_l) = y + \frac{R}{n_N} \sum_{j \neq l+1}^{n_N} g_j - \frac{\beta_l}{n_N - 1} \sum_{j \neq l+1}^{n_N} g_j - (1 - \frac{R}{n_N}) g_l + \beta_l \frac{n_N - l}{n_N - 1} g_l - \alpha_l \frac{l-1}{n_N - 1} g_l$$

The first three terms on the right-hand side of equation (A3) are equivalent to player l 's utility if she would deviate and contribute zero while the last three terms summarize the utility loss through contributing $g_l > 0$. Thus (A3) can be rewritten as:

$$(A4) \quad U_l(g_l) = U_l(g_l = 0) - (1 - \frac{R}{n_N}) g_l + \beta_l \frac{n_N - l}{n_N - 1} g_l - \alpha_l \frac{l-1}{n_N - 1} g_l$$

Since $\alpha_l \geq \beta_l$, $l \geq k + 1$, and $\beta_l < 1$, an upper bound for l 's utility from contributing is:

$$(A5) \quad \begin{aligned} U_l(g_l) &\leq U_l(g_l = 0) - (1 - \frac{R}{n_N}) g_l + \beta_l \frac{n_N - l}{n_N - 1} g_l - \beta_l \frac{l-1}{n_N - 1} g_l \\ &\leq U_l(g_l = 0) - (1 - \frac{R}{n_N}) g_l + \beta_l \frac{n_N - 2(k+1) + 1}{n_N - 1} g_l \\ &< U_l(g_l = 0) - (1 - \frac{R}{n_N}) g_l + \frac{n_N - 2k - 1}{n_N - 1} g_l \\ &= U_l(g_l = 0) - \frac{(1 - R/n_N)(n_N - 1) - (n_N - 2k - 1)}{n_N - 1} g_l \end{aligned}$$

Thus if

$$(A6) \quad \frac{(1 - R/n_N)(n_N - 1) - (n_N - 2k - 1)}{n_N - 1} \geq 0$$

a deviation of player l to a contribution of zero is profitable. Equivalent transformation of (A6) yields that l has an incentive to deviate to a contribution of zero if and only if

$$(A7) \quad k \geq \frac{n_N - 1}{2n_N} R.$$

However, since $0 \leq \frac{n_N - 1}{2n_N} < \frac{1}{2}$, the value $R/2$ is an upper bound for $\frac{n_N - 1}{2n_N} R$. Hence, if $k \geq R/2$ we can be sure that no equilibrium with strictly positive contributions exists. In other words, if the number of players k with $R/n_N + \beta_i < 1$, is at least $R/2$, all players contribute zero, independent of the community size n_N . If $R < 2$ (as in our experimental setting with $R = 1.6$) the presence of already one single player with $R/n_N + \beta_i < 1$ prevents cooperation in N , independent of community size n_N .

II. It remains to be shown that in case all players have $R/n_N + \beta_i \geq 1$ equilibria with strictly positive contributions exist and in all these equilibria, all players contribute the same amount.

First we show that under the assumption that all players in N satisfy $R/n_N + \beta_i \geq 1$ there exists a multiplicity of (pure strategy) equilibria in which all players contribute an identical amount of $g_N \in \{0 \dots y\}$ to the joint project. If all players contribute g_N , then player i has no incentive to deviate to a lower contribution $g_i < g_N$ since for her (by assumption) the monetary benefit of withholding 1 unit is lower (or equal) than the total loss from deviation $R/n_N + \beta_i$. Player i has also no incentive to deviate to a higher contribution $g_i > g_N$ since in this case a contribution increase by 1 unit would cause a strictly positive utility loss of $1 - R/n_N + \alpha_i > 0$. Thus, if all players contribute the same amount, no player has an incentive to deviate, neither to a higher nor to a lower contribution.

Are there additional equilibria with non-identical contributions? Assume that there are two different contribution levels: l players contribute $g^L > 0$ while $(n_N - l)$ players contribute $g^H > g^L$. A player j with $R/n_N + \beta_j \geq 1$ is ready to increase her contribution if *all* other players contribute more than player j since for her the monetary benefit of withholding one monetary unit is lower (or equal) than her total loss in utility from deviation $R/n_N + \beta_j$.

Thus, no player has an incentive to contribute less than g^L . An analogous argument shows that no player has an incentive to contribute more than g^H .

The situation with l players contributing g^L and $(n_N - l)$ players contributing g^H can only be an equilibrium if the players contributing g^L have no incentive to increase their contributions towards g^H while the players contributing g^H should not have an incentive to decrease their contributions towards g^L . In the following, we deduce the condition which has to be satisfied such that players with contributions g^H have no incentive to deviate to a lower contribution (Part A). Moreover, we deduce the condition which has to be satisfied such that players with contributions g^L have no incentive to deviate to a higher contribution (Part B).

Part A. If a high contributor i deviates from g^H by reducing her contribution by Δ , with $g^H - \Delta > g^L$, then her utility increases by $\Delta - \frac{R}{n_N} \Delta + \frac{\alpha_i}{n_N - 1} l \Delta - \frac{\beta_i}{n_N - 1} (n_N - l - 1) \Delta$. If this term is negative then a deviation to a lower contribution is not profitable. The term is negative if and only if

$$(A8) \quad l \leq \frac{(R/n_N + \beta_i - 1)(n_N - 1)}{\alpha_i + \beta_i}.$$

This means that a high contributor does not decrease her contribution if the number of low contributors is not too high (cf. Fehr and Schmidt 1999, proof of Proposition 4 part c), p. 862).

Part B. If a low contributor i deviates from g^L by increasing her contribution by Δ , with $g^H - \Delta > g^L$, then she gains $-\Delta + \frac{R}{n_N} \Delta - \frac{\alpha_i}{n_N - 1} (l - 1) \Delta + \frac{\beta_i}{n_N - 1} (n_N - l) \Delta$.

If this term is negative then a deviation is not profitable. The term is negative if and only if

$$(A9) \quad l \geq \frac{n_N(R/n_N + \beta_i - 1) + 1 - R/n_N + \alpha_i}{\alpha_i + \beta_i}.$$

Thus, a low contributor does not increase her contribution if the number of low contributors is

not too low.

Hence, strategy combinations with two groups of conditional cooperators contributing different amounts can only be part of an equilibrium if l satisfies both conditions (A8) and (A9). In the following we show that there is no such l .

$$(A10) \quad \Leftrightarrow \frac{n_N(R/n_N + \beta_i - 1) + 1 - R/n_N + \alpha_i}{\alpha_i + \beta_i} \leq l \leq \frac{(R/n_N + \beta_i - 1)(n_N - 1)}{\alpha_i + \beta_i}$$

$$\Leftrightarrow \frac{n_N(R/n_N + \beta_i - 1)}{\alpha_i + \beta_i} + \frac{1 - R/n_N + \alpha_i}{\alpha_i + \beta_i} \leq l \leq \frac{(n_N - 1)(R/n_N + \beta_i - 1)}{\alpha_i + \beta_i}$$

Because all players are assumed to have $R/n_N + \beta_i - 1 \geq 0$ and $1 - R/n_N > 0$ is satisfied in each public goods game, the lower bound of l is strictly greater than the upper bound of l . Hence, there is no l that satisfies (A10). This means that the situation with two groups of conditional cooperators who contribute different amounts cannot be part of an equilibrium. **Q.E.D.**

Adaption of the Fehr and Schmidt (1999) results to the punishment community (cf. Fehr and Schmidt 1999, Proposition 5, p. 841):

Proposition 2. Assume there are two types of players, n_p' “conditional cooperative enforcers” (short: “enforcers”) with preferences that obey $R/n_p + \beta_i \geq 1$ and $c < \frac{\alpha_i}{(n_p - 1)(1 + \alpha_i) - (n_p' - 1)(\alpha_i + \beta_i)}$ for $i \in \{1, \dots, n_p'\}$ and $(n_p - n_p')$ players who are only interested in their own monetary payoff (short: “payoff maximizers”), i.e., $\alpha_j = \beta_j = 0$ for all $j \in \{n_p' + 1, \dots, n_p\}$. Then the following actions are part of a subgame perfect equilibrium in P:

- I. On the equilibrium path all players contribute the same amount $g \in [0, 20]$ and no punishment occurs in the punishment stage.
- II. If, off the equilibrium path, one of the payoff maximizers chooses $g_j < g$ then each enforcer punishes the deviator with the punishment level $t_{ij} = (g - g_j)/(n_p' - c)$ while all other players do not punish.

Proof. By following backward induction reasoning, we first consider the punishment stage.

Suppose that one of the payoff maximizers $j \in \{n_p'+1, \dots, n_p\}$ deviates and chooses $g_j < g$. We show that this deviator is punished by all enforcers with the punishment level t_{ij} as stated above and that this makes the deviation not profitable.

If in the punishment stage all enforcers choose a punishment level $t_{ij} = (g - g_j)/(n_p' - c)$, then deviator j obtains the same monetary payoff as each enforcer $i \in \{1, \dots, n_p'\}$. To see this consider the monetary payoffs of j and i :

$$(A11) \quad x_j = y - g_j + \frac{R}{n_p} [(n_p - 1)g + g_j] - n' \frac{g - g_j}{n' - c}$$

$$(A12) \quad x_i = y - g + \frac{R}{n_p} [(n_p - 1)g + g_j] - c \frac{g - g_j}{n_p' - c} - \frac{n_p' - c}{n_p' - c} (g_j - g_j)$$

The right hand side of (A12) can be rewritten as

$$(A13) \quad y - g_j + \frac{R}{n_p} [(n_p - 1)g + g_j] - (n_p' + c - c) \frac{g - g_j}{n_p' - c} = x_j$$

which demonstrates that the deviating payoff maximizer achieves the same payoff as each enforcer. This payoff, however, is strictly lower than the payoff of a payoff-maximizer who did not deviate and contributed g . Thus, given t_{ij} a deviation of a payoff maximizer to a lower contribution is not profitable. Obviously, a deviation to a higher contribution level is also not profitable.

Now, we have to assure that enforcers' punishment strategies are credible, i.e., that an enforcer does not have an incentive to unilaterally reduce her punishment t_{ij} . If an enforcer reduces t_{ij} by δ she saves δc and experiences less disadvantageous inequality relative to the $(n_p - n_p' - 1)$ non-enforcers. This creates a utility gain of $[\alpha_i (n_p - n_p' - 1) \delta c] / (n_p - 1)$. On the other hand, the enforcer also experiences disutility from the disadvantageous inequality with respect to the defector j and advantageous inequality with respect to the other $(n_p - 1)$ enforcers who stick to the punishment t_{ij} . The disadvantageous inequality causes a utility loss of $\alpha_i (1 - c) \delta / (n_p - 1)$ whereas the advantageous inequality reduces the utility by $\beta_i (n_p' - 1) \delta \varepsilon / (n_p - 1)$. Thus the total utility loss from a reduction in t_{ij} is greater than the gain

if

$$(A14) \quad \frac{1}{n_p - 1} [\alpha_i(1-c)\delta + \beta_i(n_p' - 1)\delta c] > \delta c + \alpha_i(n_p - n_p' - 1) \frac{\delta c}{n_p - 1}$$

holds. One can easily show that (A14) is equivalent to

$$(A15) \quad c < \frac{\alpha_i}{(n_p - 1)(1 + \alpha_i) - (n_p' - 1)(\alpha_i + \beta_i)},$$

i.e., the condition we assumed in the Proposition. Obviously, a deviation to a higher punishment level is also not profitable. It would cause a monetary loss, disadvantageous inequality with respect to the other enforcers, would increase the disadvantageous inequality with respect to the non-punishing contributors and would cause advantageous inequality with respect to the punished player. Hence, the punishment level $t_{ij} = (g - g_j)/(n_p' - c)$ provides no incentives for deviation and is thus credible.

Do enforcers have an incentive to deviate in the contribution stage? Suppose the deviating enforcer reduces her contribution by $\delta > 0$. The deviator i gains $(1 - R/n_p)\delta$ in monetary terms but she experiences a disutility of $\beta_i\delta$ from the advantageous inequality with respect to all other players. Since, by assumption, $1 - R/n_p \leq \beta_i$ and since the player may additionally experience punishment in stage 2, this deviation does not pay. Hence, no enforcer deviates in the contribution stage either. On the other hand, choosing $g_i > g$ is not profitable for any player either, since it reduces the monetary payoff and increases inequality. **Q.E.D.**

Corollary 1. For the class of equilibria described in Proposition 2, enforcers can only exist if $(n_p - n_p') < 1/c$, i.e., the number of payoff maximizers is strictly lower than the reciprocal value of the cost of punishing.

Proof. Assume the existence of n_p' enforcers in P who satisfy (A15) which can be rephrased as

$$(A16) \quad c(n_p - n_p') < 1 - \frac{c(n_p - 1) - c\beta_i(n_p' - 1)}{\alpha_i}.$$

By contradiction we show that $(n_p - n_p') < 1/c$ has to be satisfied. Assume $(n_p - n_p') \geq 1/c$.

Then (A15) would imply that $\frac{c(n_p - 1) - c\beta_i(n_p' - 1)}{\alpha_i} < 0$. This, however, can never be the case because $\frac{c(n_p - 1) - c\beta_i(n_p' - 1)}{\alpha_i}$ is always strictly positive for each of the n' enforcers.

Hence $(n_p - n_p') < 1/c$ has to be satisfied. **Q.E.D.**

An intuition for this potentially unexpected implication is the following: 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 $[\alpha_i / (n_p - 1)](1 - c)$. At the same time, the enforcer creates a payoff inequality of c units with respect to each non-enforcer who contributes but does not punish. This means that the enforcer suffers from a disutility $[\alpha_i / (n_p - 1)]c$ with respect to each non-enforcer; in sum $[\alpha_i(n_p - n_p' - 1) / (n_p - 1)]c$. 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., $[\alpha_i / (n_p - 1)](1 - c) > [\alpha_i(n_p - n_p' - 1) / (n_p - 1)]c$. This condition is equivalent to what Corollary 1 proposes: $(n_p - n_p') < 1/c$.

Corollary 1 implies for $c = 1/3$ that in an equilibrium of the class above the punishment community can "afford" at most two payoff-maximizers independent of the community size n_p . If there are no enforcers who threaten to punish the non-contributors, the situation is "equivalent" to the situation in N described above: equilibria with positive contributions only exist if all players are conditional cooperators with $R/n_p + \beta_i \geq 1$, independent of n_p .

Analysis of the reward community on base of the Fehr and Schmidt (1999) results:

Lemma 1. 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.

Proof. Assume that each of the n_s' players rewards every other contributor with the same reward amount ρ^S . If the costs of rewarding are $c = 1$ inequality cannot be reduced. Thus, 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.

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 Proposition 1 part I, 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 3 shows that this is not the case.

Proposition 3. 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 3 is as follows: We have shown that players who contribute to the joint project may equalize their payoffs to free-riders' payoffs (Lemma 1). 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.

Proof. 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 3. 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:

$$(A17) \quad u_i^D = y - g + a_s(n_s' + 1)g + n_s' \rho^S - \left(\frac{\beta_i}{n_s - 1} (n_s - n_s' - 1)(n_s' \rho^S - g) \right) - \left(\frac{\beta_i}{n_s - 1} n_s' \rho^S (1 + n_s' c) \right)$$

In the first line of (A17) the monetary payoff of the deviating free-rider is shown; the first 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 second term in the second line (in big parentheses) 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 (A17) 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 :

$$(A18) \quad \beta_i < \frac{(n_s - 1)[a_s + c(1 - a_s)]}{(n_s - 1)c + 1}$$

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 (A18) 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 (A18)

thus having always an incentive to deviate. Hence, a separating equilibrium as described in Proposition 3 cannot exist.

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?

Lemma 2. 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

$$(A19) \quad c < \frac{\beta_j(n_s'-1) - \alpha_j(n_s - n_s')}{(n_s - 1)[n_s - 1 - \beta_j(n_s'-1) + \alpha_j(n_s - n_s')]}$$

holds for each of the n_s' players. A conditional cooperator who satisfies (A19) is called a *reward provider*.

Lemma 2 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 (A19) we may observe an equilibrium in which all players contribute $g > 0$ to the joint project and reward providers reward each contributor?

Proposition 4. Consider a community of n_s players with $n_s' \geq 2$ reward providers. Then there exists a pooling equilibrium (in contributions) in the reward community 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 4 is as follows: In Proposition 3 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 4 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.

Proof. 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:

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

The utility of the reward deviator is then:

$$(A20) \quad u_j^D = y - g + a_s n_s g - (n_s - 1)c\rho^D + (n_s' - 1)\rho^* \\ - \left(\frac{\alpha_j}{n_s - 1} (n_s - n_s') \rho^D [1 + (n_s - 1)c] \right) - \left(\frac{\beta_j}{n_s' - 1} (n_s' - 1) (\rho^* - \rho^D) [1 + (n_s - 1)c] \right)$$

In the first line of (A20) the monetary payoff of the reward deviator is shown; the first 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 second term in the second line (in big parantheses) 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 (A20) 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

$$c < \frac{\beta_j (n_s' - 1) - \alpha_j (n_s - n_s')}{(n_s - 1)[n_s - 1 - \beta_j (n_s' - 1) + \alpha_j (n_s - n_s')]} . \text{ Q.E.D.}$$

Calculations to Proposition 4. The derivative of (A20) with respect to ρ^D is:

$$(A21) \quad \frac{\partial u_j^D}{\partial \rho^D} = -(n_s - 1)c - \frac{\alpha_j}{n_s - 1}(n_s - n_{s'})[1 + (n_s - 1)c] \\ + \frac{\beta_j}{n_s - 1}(n_{s'} - 1)[1 + (n_s - 1)c]$$

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 (A21) 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 (A19), i.e., $c < \frac{\beta_j(n_{s'} - 1) - \alpha_j(n_s - n_{s'})}{(n_s - 1)[n_s - 1 - \beta_j(n_{s'} - 1) + \alpha_j(n_s - n_{s'})]}$.

Corollary 2 (a) If all players in the reward community are conditional cooperators and $n_s \geq 3$, then (A19) is satisfied. This means that there exists an equilibrium as described in Proposition 4. **(b)** If the reward community 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 4.

Corollary 2 (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 2 states that in the presence of free-riders for $n_s \leq 3$ an equilibrium as described in Proposition 4 never exists.

Corollary 3. Consider a community of n_s players. Suppose that there are no reward

providers satisfying (A19), 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 the reward community is analogous to the situation in N. This means, in particular, if there is a sufficient number of free-riders in the reward community, cooperation is impossible and $g_i = 0$ is the unique equilibrium.