Dannenberg, Astrid; Haita-Falah, Corina; Zitzelsberger, Sonja

Working Paper

Voting on the threat of exclusion in a public goods experiment

MAGKS Joint Discussion Paper Series in Economics, No. 08-2019

Provided in Cooperation with:
Faculty of Business Administration and Economics, University of Marburg

Suggested Citation: Dannenberg, Astrid; Haita-Falah, Corina; Zitzelsberger, Sonja (2019) : Voting on the threat of exclusion in a public goods experiment, MAGKS Joint Discussion Paper Series in Economics, No. 08-2019, Philipps-University Marburg, School of Business and Economics, Marburg

This Version is available at:
https://hdl.handle.net/10419/204803

Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 08-2019

Astrid Dannenberg, Corina Haita-Falah and Sonja Zitzelsberger

Voting on the Threat of Exclusion in a Public Goods Experiment

This paper can be downloaded from
http://www.uni-marburg.de/fb02/makro/forschung/magkspapers

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Voting on the Threat of Exclusion in a Public Goods Experiment

Astrid Dannenberg[a]*, Corina Haita-Falah[a], and Sonja Zitzelsberger[a]

[a] Department of Economics, University of Kassel, 34117 Kassel, Germany. Emails: dannenberg@uni-kassel.de, corina.haita@gmail.com, sonja.zitzelsberger@uni-kassel.de.

* Corresponding author

Abstract

Ostracism is practiced by virtually all societies around the world as a means of enforcing cooperation and excluding members who show anti-social behaviors or attitudes. In this paper, we use a public goods experiment to study whether groups choose to implement an institution that allows for the exclusion of members. We distinguish between a costless exclusion institution and a costly exclusion institution that, if chosen, reduces the endowment of all players. We also provide a comparison with an exclusion institution that is exogenously imposed upon groups. A significant share of the experimental groups choose the exclusion institution, even when it comes at a cost, and the support for the institution increases over time. Average contributions to the public good are significantly higher when the exclusion option is available, not only because low contributors are excluded but also because high contributors sustain a higher cooperation level under the exclusion institution. Subjects who vote in favor of the exclusion institution contribute more than those who vote against it, but only when the institution is implemented. These results are largely inconsistent with standard economic theory but can be better explained by assuming heterogeneous groups in which some players have selfish and others have social preferences.

Keywords: public goods experiment; cooperation; ostracism; institutional choice; social preferences

JEL Classification: C72, C91, C92, D02, D71, H41

1. Introduction

Humans are said to be unique in their ability to cooperate (Bowles and Gintis, 2003). Cooperation among nonrelatives occurs frequently, for example among employees, members of a work team, or users of a common pool resource. Stable cooperation often relies on actual or potential punishment of defectors. Punishment can take various forms, ranging from soft measures like disapproval to material measures like fines to harsh punishment like ostracism. Punishment may be assigned and enforced by an external authority, for example by the nation state or the employer, or it may be initiated and enforced within the community. Numerous studies in the lab and in various field contexts have shown that people are willing to punish defectors even at a personal cost (Ostrom, 1990; Chaudhuri, 2011). This wide-spread willingness to punish defectors allows communities to establish and maintain cooperation even without intervention by an external authority. When giving the choice between solving a cooperation problem without the possibility to punish others and a setting with such a possibility, people initially appear to be skeptical of the idea of punishment but often come to appreciate it when they gain experience (Gürerk et al., 2006; Dannenberg and Gallier, 2019).

In this paper, we investigate a particular form of punishment, namely ostracism, in an experimental setting. Our main interest is on whether people choose ostracism as a punishment institution when they have the choice and how this decision affects cooperation within the group. For reasons of comparisons, we also consider the effects of ostracism when the institution is exogenously given.

Ostracism refers to the general process of excluding individuals from a group. It has been practiced in virtually all societies throughout all recorded history, from ancient Rome and medieval European kingdoms, to traditionalist communities like the Amish or clans in Tribal Montenegro, to modern Western democratic societies (Boehm, 1986; Gruter, 1986; Zippelius, 1986). Ostracism in various forms is embedded in our legal traditions and it is used in the formal and informal legal procedures to increase social cohesion. Formal procedures such as imprisonment can be described as the modern transformation of traditional forms of ostracism, executed and enforced by governmental institutions (Gruter and Masters, 1986). Many groups that exist in modern societies, like political parties, companies, universities, or nonprofit associations, have rules that determine if and under which circumstances a member can be excluded. These exclusion rules may be implemented fully at the group's own discretion or they may be restricted by superior regulations. For example, unions are typically not allowed to exclude individuals from the negotiated improvements of the working

conditions (Traxler et al., 2001). Universities award tenure to their faculty members but there is limited scope to withdraw the decision later. On the other hand, political parties and nonprofit associations usually have discretion in using and determining exclusion rules and they often allow for exclusion if members violate important principles (Bolleyer and Gauja, 2015). For example, the statute of the Alliance 90/The Greens in Germany contains the following statement: "A member who willfully violates the statute or substantially violates principles of Alliance 90/The Greens and by this causes serious harm to the party can be excluded."[1] The statute of the European People's Party states "The suspension and the exclusion of a member may only be decided by the Political Assembly. It is not obliged to disclose its reasons."[2] Users of common pool resources implement exclusion rules, among other things, to secure a sustainable use of the resource. For example, small villages in Switzerland and Japan have established rules for managing communal land as well as measures for violations of the rules including, as the ultimate punishment, banishment from the village (Ostrom, 1990). Microfinance groups whose members borrow under joint liability often exclude individuals who fail to repay the loan from non-credit activities such as helping in organizing family events or crops collection (Baland et al., 2017; Putnam et al., 1994). Exclusion is also relevant for the digital commons. The designers of "Gnutella," a large peer-to-peer file sharing network, decided to offer their users the option to prevent individuals from downloading their files if these individuals themselves do not share their files in return (Strahilevitz, 2003).

Unlike the deprivation or impairment of property (monetary punishment), ostracism necessarily is a collective decision as it requires some form of coordinated response by the community members. This can explain why it has been predominantly used for crimes that affected the community as a whole, such as cultic violations, arson, or high treason (Zippelius, 1986). The right to social participation is contingent on an individual's ability and willingness to achieve and maintain an acceptable level of cooperation within the community. The immediate consequence of excluding non-cooperative individuals is that the society becomes smaller. The indirect and longer-term consequence is that further decline of cooperation may be averted. Ostracism can be useful in supporting group cohesion but it can also hurt the community if too many or the wrong individuals are excluded (Gruter and Masters, 1986), a risk that appears to be particularly dangerous for groups

[1] BÜNDNIS 90/DIE GRÜNEN (2016), „Grüne Regeln," Berlin, §21, Art. (3), available at www.gruene.de/satzung (accessed January 2019).
[2] Statutes of the European People's Party (2018), available at https://www.epp.eu/files/uploads/2019/01/EPP-Statutes-adopted-by-the-Helsinki-Congress-on-7-Nov-2018.pdf (accessed January 2019).

that depend on sufficiently large membership to achieve their goals. Even if ostracism is exclusively targeted at defectors, the unforgiving nature of the punishment may preclude potential rehabilitation and, together with the provisions that may be needed to separate the excluded members from the group, make the punishment overly expensive.

Despite the widespread incidence of ostracism in human societies around the world, the economics literature has devoted only little attention to the phenomenon, especially when compared to the study of monetary punishment, which has received considerable attention (Ostrom et al., 1992; Fehr and Gächter, 2000; for a review see Chaudhuri, 2011). The existing studies of ostracism (Masclet, 2003; Cinyabuguma et al., 2005; Maier-Rigaud et al., 2010; Akpalu and Martinsson, 2011) show that an exogenously given option to exclude members from the group increases cooperation compared to the situation without this option (see the next section for an overview of the related literature).

In this paper, we use a repeated linear public goods game to study whether groups choose to implement an exclusion institution when they have a choice.[3] Depending on their choice, groups either have the option to exclude members over the course of the game or not. Excluded players still receive the endowment but they are no longer able to contribute to the public good, nor do they benefit from it. We distinguish between a costless exclusion institution and a costly exclusion institution which, if chosen, reduces the endowment of all players. In all conditions, the social optimum that maximizes the collective payoff of the group is reached if all players contribute their full endowment to the public good. Hence, excluding a player necessarily means that the social optimum is no longer available as the group loses a potential contributor. If there is an institutional cost to the exclusion option, then implementing the institution forecloses achievement of the social optimum even if no group member is excluded. With this design, we want to test if the experimental groups choose to implement the exclusion institution, how this choice affects cooperation and efficiency, if there are systematic behavioral differences between supporters and opponents of the institution, and how a fixed institutional cost affects the decisions and outcomes. We also compare an exclusion institution that is endogenously chosen by the groups themselves to one that is

[3] Following Masclet (2003), Cinyabuguma et al. (2005), and Maier-Rigaud et al. (2010), we describe our game as "public goods game," even though players may be excluded from consuming the good. It is clear, of course, that the exclusion of members is infeasible for public goods like national defense, enforcement of law, or clean air.

exogenously imposed to understand the robustness of the results with respect to how the institution is implemented.

The experimental design clearly represents a marked simplification of the institution formation process in the real world which usually is a slow process with gradual changes over time. In many of the above-mentioned examples of ostracism, especially for those states of affairs that have a long history, it is impossible to say when exactly members agreed to use ostracism as a way to punish wrong-doers. In some cases, exclusion might have started as an ad-hoc reaction by a few members of the society and then developed into a social norm or tradition without ever being openly discussed and democratically chosen. Nevertheless, at any given point in time, the functioning of an institution depends on its acceptance and the emergence or preservation of an institution is a consequence of an internal agreement of (at least some of) the society's members. The treatment of wrongdoers represents an important choice for a society and it is often at least implicitly included in the political competition and voting decisions. Curtailing the institution formation process into a limited number of decisions in a short period of time allows us to study the preferences for the exclusion institution in a highly controlled setting, to compare the performance of groups that implement the institution and groups that do not implement it, and to compare the behavior of the supporters and the opponents of the institution.

We find that a significant share of the experimental groups choose to implement the exclusion institution. The institution is chosen less often when it comes at a fixed cost than when it is costless, but either way the support for the institution increases over time when players gain experience. Groups exclude on average one group member (out of five) and this is always the lowest contributor. Contributions to the public good are significantly higher when the exclusion option is available, not only because low contributors are excluded but also because high contributors sustain a higher cooperation level under the exclusion institution. Supporters of the institution contribute more than its opponents when the institution is implemented, while there is no significant difference between supporters and opponents when the institution is not implemented. With respect to how the institution is implemented, we find that groups that choose the institution contribute slightly more than groups that are forced to play under the same institution. The differences, however, are small and not statistically significant.

These results are to a large extent inconsistent with the standard economics model based on purely selfish preferences. According to the standard economics model, the exclusion institution does not

alter the zero-contribution equilibrium, as the threat of exclusion is not sufficient to support cooperation in a finitely repeated game. Given the inconsistency between the standard economic model and the experimental results, we use two simple and well-established models to show that the results can be better explained by assuming social preferences. The inequality aversion model by Fehr and Schmidt (1999) assumes that individuals dislike income differences between themselves and others. The reciprocity model by Rabin (1993) assumes that individuals derive utility from repaying kindness with kindness and unkindness with unkindness. The two models make similar predictions for the choice of the institution and the experimental results closely resemble the predictions for heterogeneous groups in which the majority of players are social and the minority is selfish. In the experiment, cooperators only profit from the exclusion institution when it is costless. When there is an institutional cost, cooperators on average earn slightly less with the exclusion institution than without it. The observation that many cooperators still vote in favor of the costly exclusion institution indicates that they not only derive utility from material payoffs but also from a more just outcome.

Our findings help to better understand the role of social preferences in the institution formation process and the regulation of social life in general. They thereby add to the growing literature suggesting that human preferences are heterogeneous and affect not only individual behavior under specific circumstances, but also the collective choice of institutions. We also contribute to the recent literature that looks at the differences between endogenously chosen institutions and exogenously imposed institutions, shedding light on the question if an institution by itself changes behavior or if it merely acts as a sorting or signaling mechanism.

The remainder of the paper is structured as follows. Section 2 provides an overview of the previous literature on cooperation in finitely repeated games, the effects of punishment opportunities, and endogenous institutional choice. Section 3 describes our experimental design and Section 4 discusses the institutional choice based on standard economic theory and two models of social preferences. Section 5 presents the main experimental results (less important results are presented in an Appendix) and Section 6 discusses the results and concludes.

2. Previous experimental literature

Numerous experiments have shown that monetary punishments increase contributions in finitely repeated public goods games (Fehr and Gächter, 2000; for a review see Chaudhuri, 2011). Punishments are typically used by conditional cooperators to punish low contributors and their magnitude is greater the more a player's contribution falls short of the others' average. Despite higher contributions, however, payoffs do not necessarily increase due to the costs that occur on the side of punishers and their targets. Clear payoff advantages are often realized only towards the end of the game or in games with particularly long time horizons (Gächter et al., 2008). Compared with monetary punishment, only few studies have investigated the effects of an exclusion institution. These studies typically include an additional stage after the contribution stage in which players are informed about individual contributions and then can vote to exclude one or more of the other players from the game for all or some of the remaining periods (Masclet, 2003; Cinyabuguma et al., 2005; Maier-Rigaud et al., 2010; Akpalu and Martinsson, 2011).[4] These studies show that subjects use the exclusion option to exclude low contributors from the group and sustain high levels of cooperation among the remaining players. Feinberg et al. (2014) show that cooperators also exclude low contributors from the group when this information is not based on their own experience but on a "gossip" note from the co-players of that low contributor in a previous game. Croson et al. (2015) show that an automatic exclusion institution that always excludes the lowest contributor leads to very high cooperation levels. Davis and Johnson (2015) study an institution in which players cannot exclude others from the benefits of cooperation but from an accompanying social activity, namely chatting with the other players. They find that players use this exclusion mechanism to punish free-riders but the overall effect of this rather soft exclusion mechanism on cooperation is small.

A number of related studies do not only look at exclusion of individual players but more broadly at sorting mechanisms that allow players to influence with whom they are playing, for instance, by letting them choose the group, switch between groups, or form new groups (Ehrhart and Keser, 1999; Page et al., 2005; Brekke et al., 2011; Charness and Yang, 2014). These experiments show that, if the available sorting mechanism allows the conditional cooperators to separate themselves from the free-riders, they often achieve much higher cooperation rates than in fixed groups and also provide an incentive for the free-riders to change their strategy.

[4] Kopányi-Peuker et al. (2018) study the effects of an exclusion institution in a weakest-link game.

Recent studies on endogenous institutional choice investigate if subjects can anticipate the positive effect of punishment on cooperation and vote in favor of a punishment institution when they have the choice. This literature distinguishes between "voting with feet," where subjects choose an institution and then play with the individuals who have chosen equally, and "majority voting," where individuals vote on an institution that, if chosen, is binding for the entire group. The literature also distinguishes between centralized institutions, which change the game's payoff structure for all players in the game, and decentralized institutions, which may or may not be used by the players to change other players' payoffs (for a review see Dannenberg and Gallier, 2019). A general result of this literature is that subjects initially are reluctant to vote for a punishment institution but learn to use it as an enforcement mechanism over time. It seems necessary, however, that imposing punishments on others is not too expensive, that voters get sufficient feedback on behavior under the different institutions, and that they can vote repeatedly (e.g. Gürerk et al., 2006; Ertan et al., 2009; Sutter et al., 2010). Strong institutions that change the nature of the cooperation game by making full cooperation one or the unique equilibrium of the game often have large effects on cooperation and are supported by many voters, at least after some rounds of learning (e.g. Tyran and Feld, 2006; Dal Bó et al., 2010; Dal Bó et al., 2018). But also weak institutions that do not change the nature of the game can have significant effects on cooperation and be quite popular (Tyran and Feld, 2002; Fehr and Williams, 2017). Institutional costs often reduce the support even though costly institutions may still be worthwhile implementing (Markussen et al., 2014; Barrett and Dannenberg, 2017).

To the best of our knowledge, it has not yet been studied how players vote when the choice is between a standard public goods game and a game with an exclusion option. In the experiment by Solda and Villeval (2018), the exclusion institution itself is exogenously imposed but players can vote to decide who will be excluded and for how long. They find that free-riders, and in particular those who deviate considerably and repeatedly from the group average, are excluded more often and for a longer period than others. They also find that players who have been excluded for a longer period are more likely to retaliate later than players who have been excluded for a shorter period.

The literature on endogenous institutions also tries to answer the question if endogenously implemented institutions have different effects on behavior than exogenously imposed institutions. A relatively robust result is that groups that implement an institution endogenously have higher cooperation rates than groups that are forced to play under the same institution. This difference

tends to be small for strong institutions, simply because strong institutions have a large effect on cooperation irrespective of how they are implemented, while the difference can be quite large for weak institutions (for a review see Dannenberg and Gallier, 2019). For example, Tyran and Feld (2006) study centralized punishment institutions in a public goods game which, if implemented, change the game's payoff structure by reducing the free-rider payoffs. They compare a strong punishment institution that makes cooperation the dominant strategy for all players and a weak punishment institution that maintains free-riding as the dominant strategy. For the strong institution, they find an average contribution rate of 96 percent when the institution is chosen by the players and 93 percent when it is exogenously imposed. In case of the weak institution, the average contribution rate is 64 percent when the players themselves have implemented the institution and it is 38 percent when the same scheme is exogenously imposed. We contribute to this literature by comparing the effects of an exclusion institution that is endogenously chosen and one that is exogenously imposed.

3. Experimental design

The public good games

Our experiment on endogenous institutions involves choosing between and playing different public goods games. The choice is always between a standard public goods game and a public goods game with an option to exclude members from the group. Participants are divided into groups of $N = 5$ members that remain fixed throughout the experiment (partner design).[5] There are four phases which consist of five rounds each, with the game being fixed within a phase. In every round, groups of size $n \leq N$ play a public goods game and every player $i \in \{1, \dots, n\}$ receives an endowment E_p of which he or she can contribute to the public good. Player $i's$ contribution is denoted by g_i. The stage game payoff to player i is given by $\pi_i = E_p - g_i + a \sum_{j=1}^{n} g_j$ and the marginal per capita return (MPCR) is $a = 0.4$.

In every round, players choose simultaneously how much to contribute to the public good. After each round, individual contributions are displayed on the screen in random order, so that it is not

[5] Studies of endogenous institutional choice typically use fixed groups to examine the emergence and development of institutions; see for example Ertan et al. (2009); Markussen et al. (2014); Kamei et al. (2015); Barrett and Dannenberg (2017).

possible to track the contribution by other members over time. This ensures that the decision to vote for the exclusion of a player in a given round is not based on player's reputation formed in the course of the game, but only on his or her contribution in that round.

To study endogenous institutional choice, we distinguish between three versions of the public goods game which are denoted by $p \in \{A, B10, B8\}$. In game A, players' endowment is $E_A = 10$. This game does not allow players to exclude other members from the group so that the group size is fixed at $n = N = 5$ in all rounds. In game B10, players' endowment is the same as in game A with $E_{B10} = 10$, whereas in game B8, it is reduced by 20 percent to $E_{B8} = 8$. Both games, B10 and B8, allow players to exclude members from the group so that $n \leq N$. For this purpose, these games include an additional stage. After having been informed about the individual contributions, players can vote to exclude a member from the group. Next to each contribution, an empty box is shown on the screen which players can tick in order to vote for that player to be excluded. The players are informed about the number of votes they have received but not from whom. Retaliation after exclusion therefore is not possible. Each player can cast at most one vote, at no cost, in order to determine who should be excluded. Players cannot vote for themselves but they can decide not to vote at all. Players who receive the votes from more than half of his or her co-players will be excluded from the game for the remaining rounds in that phase. This implies that the group can shrink over time. If the group consists of five members, a player must receive at least three votes in order to be excluded. If the group consists of three or four members, a player must receive at least two votes in order to be excluded. If the group consists of only two members, exclusion is no longer possible. With these voting rules, it is possible but unlikely that two players are excluded from the group at the same time. The only case in which two players could be excluded at the same time is when there are four players and two of them receive exactly two votes. The excluded players receive the endowment, either E_{B10} or E_{B8}, in each round but they are no longer able to contribute to and benefit from the public good. They are able to observe what happens in the public goods game but they are no longer allowed to vote for other players to be excluded.[6] There is no exclusion stage in the last round of a phase. To exclude the ostracized players from the benefits of the public good but not from getting their endowment is a relatively conservative approach. It can be

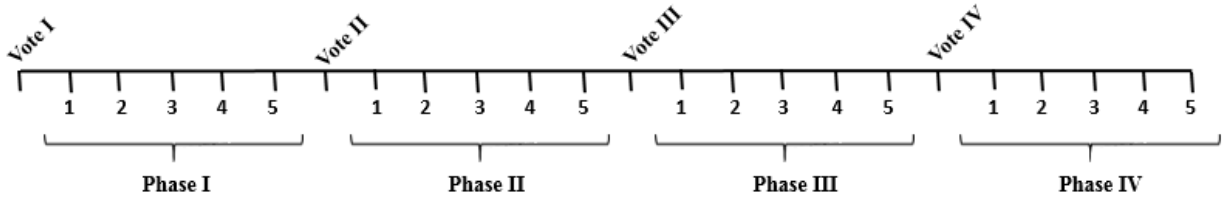
[6] The design of the B game, including the voting rules, is the same as in Maier-Rigaud et al. (2010).

interpreted that the community has the power to exclude individuals from the social benefits but not to take away their source of livelihood.

Main treatments

At the start of each phase, the full group, consisting of $N = 5$ members, chooses the game they want to play, with simple majority deciding. Importantly, the choice is always between the A game and one of the two B games (and never between the two B games). In the treatment called "*B10,*" players choose between A and B10, while in the treatment called "*B8,*" players choose between A and B8.[7] The reduced endowment in B8 compared to game A can be interpreted as a collective cost of the exclusion option. We set the fixed cost of the institution to 20 percent of the endowment, so that it would be challenging but not impossible to compensate for the cost through higher contributions in the B8 game. All members of the group simultaneously vote either for game A or for game B. There are no abstentions. For a game to be selected, at least three out of the five members must vote for that game. Members are informed about which game has been selected, but not about the individual votes. Afterwards, the group plays the chosen game throughout that phase. If the group plays B10 or B8 and a player gets excluded from the group, the exclusion lasts only until the end of the respective phase. At the beginning of the new phase, the excluded player re-enters the group and all players vote again to choose between game A and game B. Figure 1 presents the time line in the experiment.

Figure 1. Voting rounds and phases



A few things about our design are worth noting. First, players can abstain from the exclusion vote but not from the vote on the institution. There are several reasons for this. The nature of our research

[7] Since the choice is always between the A game and one of the B games, there is no mention of B10 or B8 in the experimental instructions but only game B. We used neutral language throughout, avoiding terms like "cooperation," "ostracism," or "punishment." The instructions can be found in the Appendix.

question, which is endogenous institutional choice and its consequences, requires an active game choice by the participants. Allowing for abstention from the institutional vote would have introduced behavioral issues out of our control. For example, playing game B would not necessarily imply that the majority has voted in favor of B. Another reason is to avoid practical inconvenience. Assume that all five players abstain from voting or there is a tie. A random device would have been needed then to determine which game is played, since one of the two games must be played. In this situation, the institutional choice would not have been endogenous. In the case of the exclusion vote these factors are less of an issue. The option to abstain is necessary here for situations in which all group members make equally high (or low) contributions. Second, given the MPCR of $a = 0.4$, contributing to the public good is inefficient once the group has shrunk to just two members. In this case, the collective benefit of contributing one unit to the public good is smaller than the cost ($0.8 < 1$). This could have been avoided by a higher MPCR. If, for example, the MPCR was increased to 0.6, contributing to the public good would be efficient even with two players only. However, in the initial group of five players, the full cooperative payoff would then be three times as large as the Nash payoff and thus create strong incentives to cooperate even without the exclusion institution. Alternatively, we could have restricted the voting rule in the B games by capping the number of excludable players at two but this would have facilitated the institutional choice between game A and game B. In our design, if players choose the B game their challenge is to maintain both a high cooperation level and a large enough group. Third, our groups start the experiment by choosing between the games with no prior experience. Therefore, all learning is endogenous as it depends on how groups choose and play over the course of the experiment. Experience has been shown to be critical for institutional choice, so a natural extension of our study would be to have subjects gain some experience in one or both games *before* they choose between them (Markussen et al., 2014; Barrett and Dannenberg, 2017).

Exogenous control treatments

With endogenous choice of the institution, where groups select themselves into the different games, it is not clear if the institution is successful because it attracts the most cooperative groups or because the institution changes the incentives to cooperate, regardless of whether the groups are particularly cooperative or not. In order to distinguish between the effect of self-selection and the effect of the institution, we conducted two additional treatments, *B10-exo* and *B8-exo*, in which

groups played games A and B over the same number of rounds but, unlike the groups in the endogenous treatments, these groups could not vote on the two games but had to play the game that was announced by the computer.[8] For each group in the endogenous treatments, we had one group in the exogenous treatments that played the exact same sequence of A games and B games (perfect matching groups). This means that, in each phase, the distribution of groups between the two games in the exogenous treatments is identical to the distribution in the corresponding endogenous treatment. To keep the difference to endogenous treatments to a minimum, players in the exogenous treatments were not informed about the sequence in advance but learned which game they would play only at the beginning of each phase. Apart from the missing voting stage and the way the games were chosen, everything in the exogenous treatments was identical to the endogenous treatments. The exogenous treatments also allow us to compare the results with the previous literature.

Implementation

The experimental sessions were held in a computer lab at the University of Magdeburg, Germany, using undergraduate students recruited from the general student population. In total, 460 students participated in the experiment with each one taking part in one treatment only (between-subject design). For our main treatments, we conducted eight sessions in June 2016 and assigned them randomly to *B10* and *B8*.[9] For the exogenous control treatments, we conducted ten sessions in September and November 2018 at the same computer lab and assigned them randomly to *B10-exo* and *B8-exo*.[10] For each of the four treatments, we had 23 groups that consisted of five players each. In each session, subjects were seated at linked computers (game software z-Ttree; Fischbacher 2007) and randomly divided into five-person groups. Subjects did not know the identities of their

[8] The self-selection effect under endogenous institutional choice is accompanied by two additional effects. First, an information effect arises because players learn whether the majority of group members has supported or opposed the institution and thus can draw conclusions about the cooperative nature of the group members. Second, the process of choosing the institution by itself can improve cooperation through, for example, strengthened feelings of group identity, which has been labeled democracy effect (Dal Bó, 2014; Dannenberg and Gallier, 2019). Our design does not allow us to distinguish between these three effects but only if they jointly lead to different behavior than the institution effect only.

[9] Sample characteristics (age, gender, study subject, and final school grade) do not significantly differ between *B10* and *B8* (T-test or Chi2 test, $p > 0.1$ each).

[10] As the control treatments were conducted later, we could not randomize between them and the main treatments. However, we paid careful attention that we recruited from the same subject pool and that the participants had roughly the same level of experience with experiments.

co-players, but they did know that the membership of their group remained unchanged throughout the session. The experimental instructions were handed out to the students and also read aloud to ensure common knowledge. They carefully explained both games, A and B, and included several numerical examples. Before subjects began playing the games, they had to answer a number of control questions. The control questions tested subjects' understanding of the games to ensure that they were aware of the available strategies and the implications of making different choices. The experiment began only when all participants had answered the control questions correctly. Questions during this process were answered privately. During the game, earnings were displayed in tokens. It was public knowledge that payments would be calculated by summing up the number of tokens earned over all rounds and by applying an exchange rate of €.05 per token. At the end of the experiment, the subjects were paid their earnings privately in cash.

4. Theoretical background

In this section we present theoretical predictions based on standard preferences and two models of social preferences. In each case, we first derive the equilibria in each game, A and B, and then provide predictions for players' choice between the two games. We assume common knowledge of preferences throughout the theoretical analysis. In order to save space we only present a summary of the main results. The complete analysis, including the corresponding proofs, can be found in the Appendix.

Standard preferences model

In the standard preferences model, in which players are purely self-interested, zero contribution by all players is the unique Nash equilibrium of the stage game since the MPCR to the public good is $a < 1$. This equilibrium is Pareto dominated by the outcome in which all players contribute their entire endowment as long as the group has more than two members. Using backward induction, it can be shown that the unique subgame perfect Nash equilibrium (SPNE) of the finitely repeated game is zero contribution by all players in each round, regardless of the game played. It is then clear that players are indifferent between game A and game B10, but prefer A to B8 as the former gives a higher endowment and thus a higher payoff. Hence, standard preferences predict that game B8 is never played when the choice is between A and B8, while about half the groups will play

game A and about half will play game B10 when the choice is between these two games. If game B10 is chosen, then exclusion can be part of an equilibrium since exclusion in our setting is costless and players are thus indifferent between excluding and not excluding a player from the group. Therefore, any configuration of votes and group sizes can be part of an equilibrium (see details in Appendix A.1).

Inequality aversion model

In the inequality aversion model by Fehr and Schmidt (1999), players derive utility from the material earnings resulting from the public goods game, but they also derive disutility if their earnings are higher than those of other group members (advantageous inequality aversion) or if their earnings are lower than those of the other group members (disadvantageous inequality aversion). Specifically, the inequality averse utility function assumed by Fehr and Schmidt (1999) is:

$$U_i(\pi_i) = \pi_i - \alpha_i \frac{1}{n-1}\sum_{j \neq i} \max\{(\pi_j - \pi_i), 0\} - \beta_i \frac{1}{n-1}\sum_{j \neq i} \max\{(\pi_i - \pi_j), 0\},$$

where π_i is player $i's$ material payoff from the public goods game, α_i measures the aversion to disadvantageous inequality and β_i measures the aversion to advantageous inequality, with $0 \leq \beta_i < 1$ and $\alpha_i \geq \beta_i$. Thus, players are more averse to disadvantageous than to advantageous inequality.

With this type of preferences any weakly positive contribution level $g_i = g \in [0, E_p]$, for all i can be supported as an equilibrium of the stage game if all group members are sufficiently averse to advantageous inequality, i.e. $\beta_i \geq 1 - a = 0.6$. In line with the original paper, we call these players conditional cooperators. This equilibrium exists in both games and it makes no use of the exclusion option in game B. It does, however, require coordination on a certain contribution level for which full contribution seems to be a natural focal point as it leads to the highest payoffs. By the usual backward induction argument it obtains that full contributions in each round is one SPNE of the finitely repeated game, regardless of the game played. Since the exclusion option in game B is not used, the choice between the games is governed by the contribution level on which players coordinate in each game. If there is coordination on the same contribution level across the games,

then it is clear that groups of inequality averse players are indifferent between playing A and playing B10, but they strictly prefer A to B8 (see Appendix A.2 for details).

When there is one selfish player in the group, that is a player for which $\beta_i < 1 - a = 0.6$, the unique equilibrium of the stage game is zero contribution by all players.[11] The reason for this is that zero contribution is the dominant strategy for the selfish player and, given this, it is also the best response of the remaining conditionally cooperative players, that is those player for which $\beta_j \geq 0.6$.[12] However, in the repeated game B, the conditional cooperators can make use of the exclusion institution and exclude the selfish player. It can be shown that, although in the first round all players contribute zero due to the presence of the selfish player, the conditional cooperators exclude her after this round and cooperation is restored for the remaining rounds of play. Since exclusion is not possible in game A, the only SPNE of the finitely repeated game A is for all players to contribute zero in every round. Given these equilibrium outcomes, the selfish player either strictly prefers game A regardless of the alternative, game B8 or B10 (if $\alpha_i > 0$), or she prefers A over B8 and is indifferent between A and B10 (if $\alpha_i = 0$).[13] The conditional cooperators strictly prefer B10 to A and they prefer B8 to A if they coordinate on a high enough contribution level after excluding the selfish player. Specifically, for our experimental parameters, these players should contribute more than 5 tokens for game B8 to be preferred (see Appendix A.2 for the detailed derivation of this condition).[14]

Reciprocity model

The reciprocity model developed by Rabin (1993) assumes that apart from the monetary gains, people also derive utility from reciprocation. The reciprocity term consists of two factors: one

[11] The situation with two selfish players is qualitatively not much different (see Appendix A.2).
[12] Our parameter values satisfy the condition from Proposition 4, Part b in Fehr and Schmidt (1999).
[13] Intuitively, by playing B and being excluded, as the equilibrium play of this game shows, the selfish player suffers from disadvantageous inequality because the rest of the players have higher monetary gains by cooperating from round 2 onwards. This does not happen when playing game A since everyone earns the endowment in the zero-contribution equilibrium. This is true in both treatments. Only when the disadvantageous inequality aversion parameter α is zero, is the selfish player indifferent between A and B10. For details, see Appendix A.2.
[14] For the cutting-edge contribution of 5 tokens, the conditional cooperators should, in addition, have a low enough aversion to advantageous inequality to compensate for the relatively low contributions. The anticipation that contributions are equal or less than 4 tokens would make players choose game A. That is because the gains from the public good are outweighed by the disutility from the aversion to advantageous inequality that would result from the exclusion of the selfish player in game B8.

captures the degree of kindness of the player towards her opponent and the second captures the beliefs of the player about the kindness of the opponent. We base our analysis on the multi-player extension of this model by Nyborg (2017) and define the reciprocal utility as

$$u_i = \pi_i + \beta_i R_i,$$

where π_i is the material payoff from the public good, β_i is the weight attributed to reciprocation, and R_i is the reciprocation term. We use the same measure of kindness as in Nyborg (2017) and define the reciprocation term as:

$$R_i = \frac{1}{n-1}\left(\sum_{j \neq i} \tilde{f}_{ji} + \sum_{j \neq i} f_{ij}\, \tilde{f}_{ji} \right),$$

where f_{ij} is the kindness of player i towards player j and $\tilde{f}_{ji}$ is i's belief about the kindness of j towards i. If all players have a sufficiently high concern for reciprocation, i.e. $\beta_i = \beta > 2E_p(1-a), \forall i = 1,...,n,$, then the stage game has two pure-strategy Nash equilibria: one in which all players contribute zero and one in which all players contribute their full endowment. For the SPNE in which one or the other of the stage-game equilibria is repeated every round, the symmetry of the equilibrium leaves the exclusion institution in game B unused. Hence, groups of highly reciprocal players are indifferent between playing A and B10, but prefer A to B8 due to the higher endowment (see Appendix A.3). If players are not sufficiently reciprocal, i.e. $0 < \beta \leq 2E_p(1-a),$ then zero contribution by all players is still the only equilibrium.

When there is one non-reciprocal player with $\beta_k = 0,$ but $\beta_i = \beta > 0,\ i \neq k$ the stage game again has two pure-strategy Nash equilibria. The non-reciprocal player contributes zero, regardless of what the reciprocal players do. Apart from the equilibrium in which all players contribute zero, there also is a pure-strategy equilibrium in which the reciprocal players contribute their full endowment, but only if they are highly reciprocal, i.e. $\beta > 2E_p(1-a)\frac{n-1}{n-3}$. These two types of equilibria exist both in game A and in game B. Therefore, the SPNE in game A is the repetition of either of the two pure-strategy Nash equilibria for five rounds of play. However, in game B, the SPNE that involves the full contribution equilibrium by the highly reciprocal players in the stage game, also includes the exclusion of the non-reciprocal player after the first round. If the reciprocal players

are only moderately reciprocal, i.e. $2E_p(1-a) < \beta < 2E_p(1-a)\frac{n-1}{n-3}$, then in game B there is yet a third SPNE in which all players contribute zero in the first round, the reciprocal players exclude the non-reciprocal player after this round and contribute their full endowments thereafter. Note that this third SPNE in game B requires a lower level of reciprocal concern than the one required in game A for full cooperation to emerge after the first round (compare $2E_p(1-a)$ to $2E_p(1-a)\frac{n-1}{n-3}$) Thus, the exclusion institution enables full cooperation after the first round, even if the reciprocal players are only moderately reciprocal.

For the choice between the games we assume that the reciprocal players play consistently across the two games either the zero or the full contribution equilibrium, when they exist in both games. If the zero contribution equilibrium is played, then players are indifferent between B10 and A, but strictly prefer A to B8. If the full contribution equilibrium is played, then both game B10 and game B8 are preferred by the reciprocal players. Even if the reciprocal players are only moderately reciprocal, in which case the full contribution equilibrium does not exist in game A and exists in game B only after the non-reciprocal player is excluded, they prefer B10 and B8 to game A and use the exclusion institution. The non-reciprocal player strictly prefers game A over B10 and B8, since game A allows her to benefit from the public good while defecting in all rounds (see Appendix A.3).

Differences between standard and social preferences

In summary, we have established important differences between the standard model based on purely selfish preferences and the social preferences models. In the standard preferences model, the exclusion institution does not change the zero contribution equilibrium, as the threat of exclusion is not sufficient to sustain cooperation. The predictions based on social preferences depend on the composition of the group and the ability of the social players (with strong preferences for equality or reciprocity) to coordinate towards the Pareto-superior equilibrium. Intuitively, groups consisting of social individuals only can sustain cooperation in both games. If at all, they will only choose the exclusion institution if it comes at no cost and they will not use it in equilibrium. If there is a selfish player in the group who does not care much about equality or reciprocity and if the social players coordinate successfully, they will implement the exclusion

institution, exclude the selfish player from the group, and cooperate thereafter. With some further restrictions, this is also true when the exclusion institution comes at a cost. Note that it is not our intention to test the two theories of social preferences with the experiment, as has been done for example by Blanco et al. (2011). We rather use the two theories to provide possible motivations why players may vote for and use the exclusion institution.

5. Results

We first describe how individuals voted between the two games in the endogenous treatments and how they performed depending on their choice of the game. We then describe the behavior of the players in the exogenous treatments and how it compares to the endogenous case. To keep the focus on institutional choice and its effects on cooperation, additional results are presented in the Appendix.

Voting behavior and game choice

Figure 2 shows how individuals vote over the four phases, how many groups play game A, the standard game without exclusion option, and how many groups play game B, the game with exclusion option, in each phase. The majority of individuals vote for game A in the first phase in both treatments, with the majority being particularly large when game B has a lower endowment (B8) than game A. However, the share of individuals who vote for B increases over the course of the four phases. The increase in votes for B is the largest from the first to the second phase and becomes smaller in later phases. The support for game B also increases when it has a lower endowment, but at a lower level. In the *B10* treatment, the share of B-voters increases from 37 percent in the first phase to 80 percent in the last phase and the share of groups that play B rises from 30 percent to 96 percent. In *B8*, the share of B-voters increases from 16 percent to 49 percent and the share of groups that play B increases from 4 percent to 52 percent. In each phase, groups are more likely to play game B in *B10* than in *B8* (Fisher's Exact test, $p < 0.05$ for each phase).[15]

[15] If not stated otherwise, we use two-sided tests and the number of groups per phase or the average per group and phase as unit of observation for the statistical tests.

Figure 2. Voting behavior and game choice by treatment

A closer look at the individual voting behavior shows that most individuals do not move back and forth between voting for A and voting for B, but vote relatively consistently. In both treatments, 83 percent of the individuals who start the first phase by voting for game A either keep voting for A until the end or switch to B at some point and then keep voting for B. In the *B10* treatment, 91 percent of the individuals who first vote for game B never switch to A. In the *B8* treatment, where game B is costly, a relatively large share of 56 percent consistently votes for B without switching to A. Likewise, at the group level, 75 percent of groups in B10 and 68 percent of groups in B8 that start the first phase by playing A either keep playing A or switch to B at some point without switching back. The groups that start by playing B never switch to A.

Table 2 shows regression results on the probability of voting for game B, conditional on treatment and the game played by the group in the previous phase. The best predictor of whether an individual votes for game A or game B is the voting decision in the previous phase, confirming that the preferences for the games are relatively stable over time. Another predictor is the payoff in game A when game A was played in the previous phase. The higher an individual's payoff in game A, the less likely this person is to vote for game B in the next phase. These results show that, unlike in theory where we assume common knowledge of preferences, players in the experiment must first learn about their co-players' preferences and then adjust their institutional choice accordingly. Despite the need to learn, the stability of the voting decisions over time is remarkable and thus consistent with the theory.

Table 2. Probability of voting for game B

	B10		*B8*	
	(1)	(2)	(3)	(4)
Previous game	Game A	Game B	Game A	Game B
Voted for game B in previous phase (d)	0.3343***	0.5578***	0.3852***	0.5147***
	(0.0698)	(0.0738)	(0.1180)	(0.1430)
Average payoff in previous phase	-0.0276**		-0.0457***	
	(0.0134)		(0.0080)	
for previously non-excluded subjects		0.1277		0.0560*
		(0.0113)		(0.0303)
for previously excluded subjects		-0.0033		-0.0537
		(0.0184)		(0.0525)
Average contribution (%) in previous phase	0.3121**		-0.1332	
	(0.1575)		(0.1031)	
for previously non-excluded subjects		0.3374*		-0.5370
		(0.1888)		(0.3880)
for previously excluded subjects		0.1615		0.0396
		(0.1575)		(0.3839)
Excluded in previous phase (d)		-.0359		0.0130
		(0.0665)		(0.1223)
Observations	125	220	250	95

Average marginal effects (discrete effects for binary variables) from random effects probit estimations (pooled binary probit estimations in column (2)) with standard errors in parentheses. Standard errors clustered by group. Level of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable is the probability of voting for game B. When game B was played in the previous phase, regressions include interaction terms between *average payoff in previous phase* and *excluded in previous phase* as well as between *average contribution in previous phase* and *excluded in previous phase*. Dummy indicators for phases are included. (d) indicates dummy variable.

Contributions

Table 3 gives an overview of average contributions, measured as percent of endowment, and average payoffs conditional on treatment, phase, and game. Contributions are substantially higher in game B than in game A, irrespective of treatment or phase. In *B10,* the average contribution across all phases is 41 percent in game A and 76 percent in game B. In *B8,* the average contribution is 41 percent in game A and 74 percent in game B. The differences in contributions between game A and game B within each treatment and phase are almost always statistically significant (Mann-Whitney-Wilcoxon (MWW) test, $p < 0.05$ for each treatment and phase, except phase II in *B10* where $p = 0.1009$).

Figure 3 shows how average contributions develop over time conditional on treatment and game. We see a strong end-of-phase effect in game B where the threat of exclusion dissolves and contributions drop to a similarly low level as in game A. This drop indicates that the learning process and the exclusions over the course of the game do not completely eliminate the uncertainty about the other players' preferences and the remaining players do not want to risk a too high contribution without the threat of exclusion.

Of course, higher average contributions in game B could simply result from the exclusion of low contributors. To test if the exclusions alone account for the differences between game A and game B, we compare the contributions provided by the highest contributors between the two games by leaving out the excluded players in game B and the lowest contributors in game A.[16] The contributions of the remaining players are significantly higher in game B than in game A, irrespective of treatment and phase (see Appendix Table B.1 and Figure B.1). Thus, the exclusion of low contributors alone cannot explain the higher average contributions in game B.

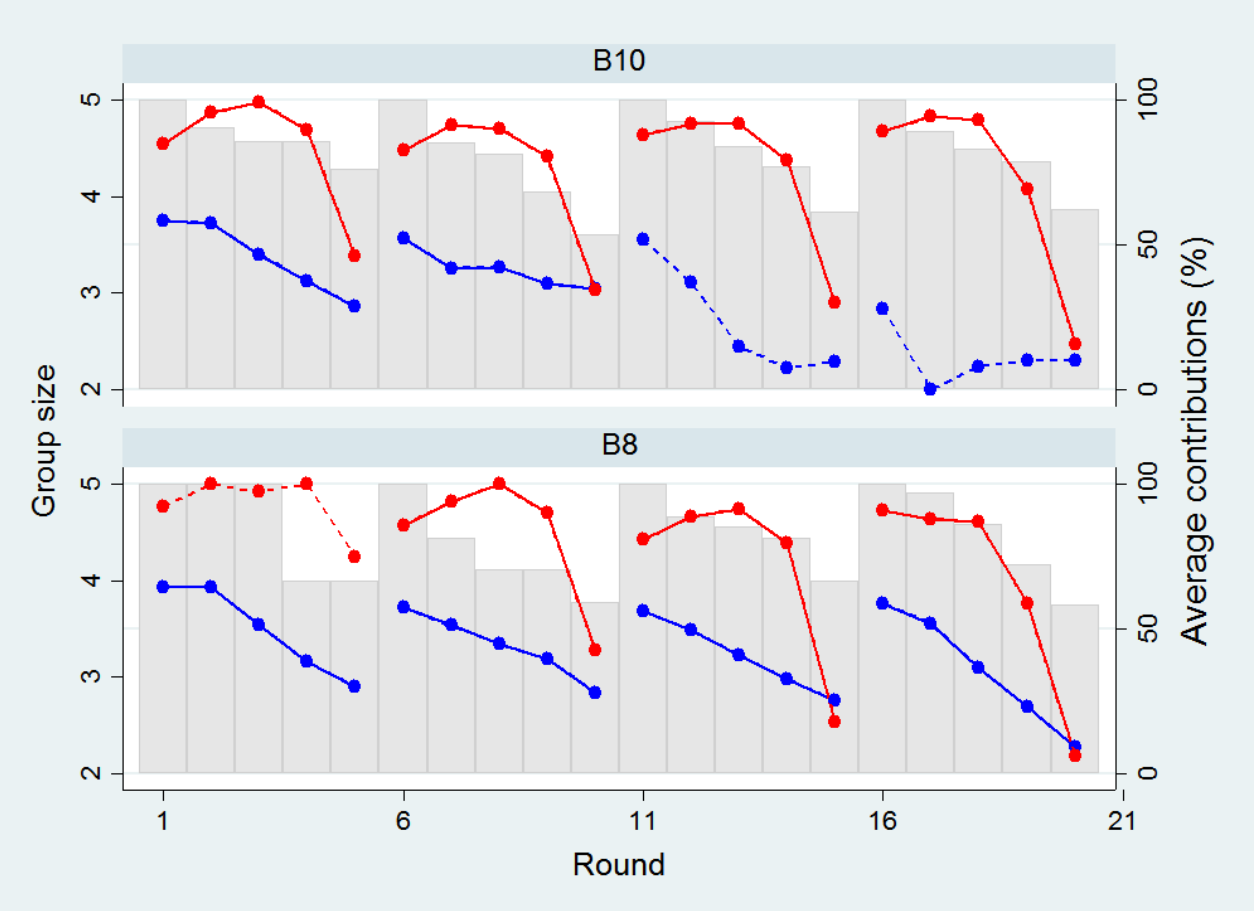
Table 3. Average contributions and payoffs by treatment, phase, and game

		B10					*B8*				
Phase	Game	Number of groups in each game	Mean group size	Average contribution (in %)	Average payoff (insider)	Average payoff (all)	Number of groups in each game	Mean group size	Average contribution (in %)	Average payoff (insider)	Average payoff (all)
I	A	16	5	45.82	14.6	14.6	22	5	49.78	15.0	15.0
	B	7	4.6	83.14***	17.4**	17.1*	1	4.6	93.00†	14.5†	14.0†
II	A	5	5	41.68	14.2	14.2	14	5	44.20	14.4	14.4
	B	18	4.3	75.91	16.4	15.8	9	4.3	82.51***	13.1	12.6
III	A	4	5	24.20	12.4	12.4	14	5	40.89	14.1	14.1
	B	19	4.5	76.19†	16.9†	16.4†	9	4.5	71.79***	13.1	12.8
IV	A	1	5	11.20	11.1	11.1	11	5	35.85%	13.6	13.6
	B	22	4.5	72.42†	16.5†	16.0†	12	4.5	66.22***	13.1	12.6
All	A		5	41.1	14.1	14.1		5	40.71	14.1	14.1
	B		4.3	75.69***	16.3***	15.7*		4.3	74.00***	12.9*	12.4***

The table shows average contributions (in percent of endowment) and payoffs conditional on treatment, phase, and game. Stars indicate statistically significant differences between groups playing game A or game B within the same treatment and phase, using a two-sided MWW test and the group average per phase as unit of observation. In the bottom two lines ("All"), stars indicate statistically significant differences within groups when they play A or B in different phases, using a Wilcoxon signed-rank test and the group's average contribution/payoff in either game across all rounds. Groups that play either A or B in all rounds are left out. Level of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. † indicates that the number of observations is too low to run a test ($N < 5$).

[16] The average group size excluding the lowest contributors in game A is four, which roughly equals the average group size in game B. In game A, in 82 percent of groups, the lowest contributor is a single player. In 2 percent of groups, all players make equally high contributions and thus have no lowest contributor. In the remaining groups, two or more players are identified as the lowest contributors.

Figure 3. Average contributions over time by treatment



The figure shows average contributions over time, measured in percent of endowment, in game A (blue) and in game B (red) by treatment. The dashed lines indicate that data points are based on only few observations (N < 5). Excluded players in game B are omitted. The lines thus represent the average efficiency level where efficiency is defined as the maximum payoff possible given the size of the groups. The bars depict average group size for groups playing game B in the respective phase.

Difference between A-voters and B-voters

In order to test if there are behavioral differences between individuals who vote for game A and individuals who vote for game B, we investigate if and how the voting decision affects the contribution decision in the same phase. We start by comparing the contribution decisions of A-voters and B-voters when they play game B for the very first time. In the *B10* treatment, A-voters contribute on average 59 percent in the first round of playing game B while B-voters contribute 83 percent in the first round. In *B8*, A-voters contribute on average 63 percent in the first round while B-voters contribute 87 percent. Table 4 provides the corresponding regression results on the differences between A-voters and B-voters when they play game B for the first time (columns (1) to (4)). The results show that having voted for game B significantly increases first round contributions in both treatments. Additional regression results, shown in the Appendix (Table B.2),

show that A-voters and B-voters do not only behave differently in the first round of playing game B, but also on average in the first phase of playing game B.

Table 4. First round contributions in game B

	Game B is played for the first time				Game B is not played for the first time			
	(1) B10	(2) B10	(3) B8	(4) B8	(5) B10	(6) B10	(7) B8	(8) B8
Voted for game B (d)	0.2294***	0.2341**	0.2509**	0.2555**	0.0561***	0.0396*	0.0817	0.0565
	(0.0651)	(0.0915)	(0.0952)	(0.0928)	(0.0217)	(0.0204)	(0.0747)	(0.0652)
Average contribution		0.3874***		0.5416***		0.1320*		0.2285
(%) in previous phase		(0.1118)		(0.0832)		(0.0702)		(0.1468)
Game B in previous						0.0495		-0.1431
phase (d)						(0.0770)		(0.1269)
Observations	115	80	75	70	215	215	80	80

OLS estimation results (Columns (1)-(4)) and random effects GLS estimation results (Columns (5)-(8)) with standard errors in parentheses. Standard errors are clustered by group. Dependent variable is the individual contribution as percentage of endowment in game B. Level of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns (5) – (8) exclude the first phase. Regressions include dummy indicators for phases. (d) indicates dummy variable.

Columns (5) to (8) in Table 4 show the differences between A-voters and B-voters when game B is played for the second, third, or fourth time. We see that having voted for B in these cases has a smaller and often insignificant effect on contributions, indicating that the differences between A-voters and B-voters wear off when they gain experience with the B-game.

We ask next if A-voters and B-voters also behave differently in game A, especially at the beginning when they have not yet gained any experience. In *B10*, when subjects play game A for the first time, A-voters contribute 56 percent in the first round while B-voters contribute 61 percent in the first round. In *B8*, A-voters contribute 66 percent in the first round and B-voters contribute 64 percent. The corresponding regression analyses on the differences between A-voters and B-voters in game A show that the voting decision only rarely affects contribution decisions in game A. Due to the mostly insignificant results, these regressions are shown in the Appendix (Tables B.3 and B.4). The regressions shown in Table 4 and in the Appendix also show that, in all treatments and games, a player's average contribution in the previous phase predicts the contribution in the current phase, indicating a relatively consistent contribution pattern over time.

Group size and exclusion of players

While the group size is fixed in game A, it is possible for groups to shrink to a minimum of two players in game B. The average group size in game B across all rounds and phases is 4.3 and the

average group size at the end of a phase is 3.8 in both treatments. In the *B10* treatment, on average across all phases, 24 percent of groups keep a group size of five throughout the phase, 47 percent reach a group size of four, 17 percent a group size of three, and 12 percent a group size of two. The respective numbers for the *B8* treatment are 26, 42, 23, and 10 percent.

Groups that do not exclude any members in game B attain very high contribution levels, which suggests that these groups happen to consist of conditional cooperators or that the mere threat of exclusion is sufficient to keep cooperation up. Groups that play game B and do exclude one or more members still achieve higher average contribution levels than groups that play game A.

The analysis of the voting decisions to exclude other players shows that many players receive a vote during the course of the game, but a much smaller share is actually excluded. Of the subjects who play game B at least once, 71 percent in *B10* and 76 percent in *B8* receive at least one vote for their exclusion. Forty-seven percent in *B10* and 37 percent in *B8* are excluded at least once over the course of the experiment. In both treatments, even though high and average contributors receive some votes, only the lowest contributors are excluded from the group. Thus, the exclusion option is used very effectively and no "anti-social" punishment occurs. Comparing the contribution behavior in game B, before and after the exclusion, we find that previously excluded subjects adapt their contribution levels in the direction of the others' average, but they still contribute less than the others. Over the same phases, non-excluded subjects keep their relative contribution levels constant and very close to the average of the others. The adjustment of the excluded players appears to be perceived as insufficient. In both treatments, we find that previously excluded individuals face a significantly higher likelihood of being excluded (again) than individuals who have not been excluded before (see Appendix Tables B.5 and B.6). Note that repeated exclusion cannot occur due to bad reputation as the contribution decisions are shown in random order in each round.

Players who receive a vote from their co-players but are not excluded can still perceive this as a warning that their contribution has been inadequate. Comparing contribution behavior before and after receiving a vote, we find that subjects who contribute less than the average of the others and who receive a vote but are not excluded adapt their contribution level in the direction of the others' average in the next round of the same phase. This is also the case for low contributors who do not receive a vote for their exclusion—but their adjustment is smaller (see Appendix Table B.7).

Payoffs

We have already established that, in both treatments, contributions in game B are significantly higher than in game A. However, this does not necessarily mean that payoffs are higher as well since the number of potential contributors in game B is lower and, in the *B8* treatment, the endowment is lower. Across all phases, we see that game B leads to slightly higher payoffs when there is no institutional cost and it leads to slightly lower payoffs when there is an institutional cost. In *B10*, the average payoff in game A is 14 tokens and the average payoff in game B is 16 tokens. In *B8*, the average payoff in game A is 14 tokens and the average payoff in game B is 12 tokens.

Table 3 (and Figure B.2 in the Appendix) show that, in *B10*, average payoffs in all phases are higher in game B than in game A and the differences is statistically significant in phase I (MWW test, $p < 0.1$). In treatment *B8*, average payoffs are lower in game B than in game A in all phases, but the differences are never statistically significant. Table B.8 in the Appendix shows that, when we compare only the highest contributors (the non-excluded players in game B and the highest contributors in game A) average payoffs are always higher in game B than in game A when there is no institutional cost, with the difference being statistically significant in phase I (MWW test, $p < 0.05$). When there is an institutional cost, the high contributors' average payoffs are lower in game B than in game A, but the difference is not statistically significant.

Comparison between endogenously chosen and exogenously imposed ostracism institution

Figure 4 provides a comparison of contribution rates in the two games between the endogenous treatments, *B10* and *B8*, and the corresponding exogenous treatments, *B10-exo* and *B8-exo*. It shows that contributions rates are very similar in the endogenous treatments and the exogenous treatments. Contributions in the B game are slightly higher in the endogenous treatments than in the exogenous treatments in both *B10* and *B8*. The same is true for the A game but only in the *B8* treatment. There is no clear tendency in the *B10* treatment. All these differences between endogenous and exogenous are very small and not statistically significant (MWW test, $p > 0.1$ each). Everything we have observed for the endogenous treatments also happens in the exogenous treatments: Contributions in the B game are significantly higher than in the A game and this is true for both *B10-exo* and *B8-exo* (MWW test, $p < 0.05$ for each treatment and phase, except phase I in *B10-exo* where $p = 0.1606$). There is a strong end-of-phase effect in the B games where contributions drop to a low level. In *B10-exo*, average payoffs are higher in the B game than in the

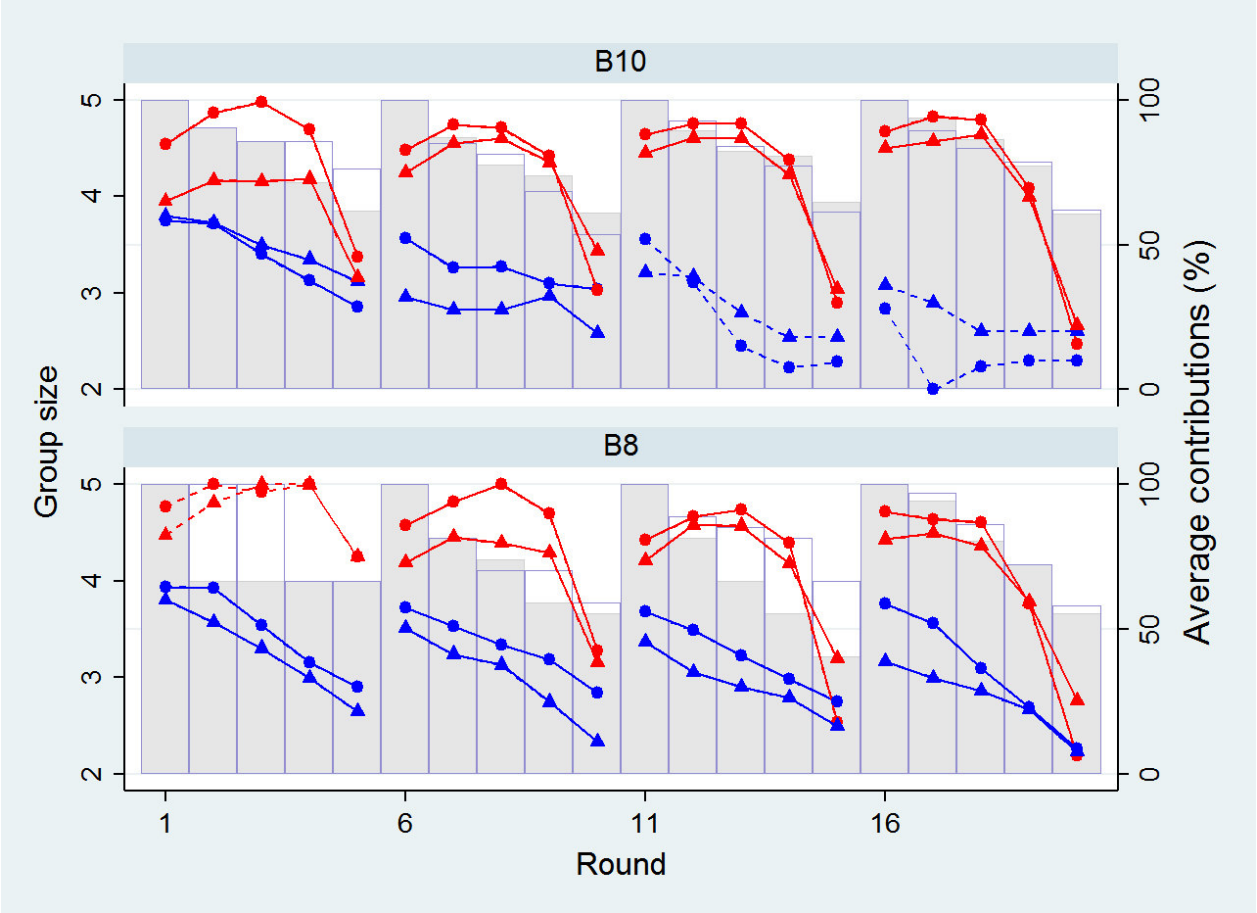
A game and the differences are significant in two phases (phases II and III, MWW test, $p < 0.1$ each). In *B8-exo*, average payoffs are higher in the A game than the B game and the differences are significant in one phase (phase II, $p < 0.1$). On average, one player is excluded in the B games and, with one exception, this is always the lowest contributor. As illustrated in Figure 4, the group size in the B game is very similar in *B10* and *B10-exo* (MWW test, $p > 0.1$ in all phases). When the exclusion institution comes at a cost, groups playing the B game tend to be slightly larger in *B8* than *B8-exo* and the difference is weakly significant in phase III ($p = 0.0965$). Regarding exclusions of individuals and exclusion proposals, we find no significant differences between the endogenous and the corresponding exogenous games ($p > 0.1$ each).[17]

Taken together, behaviors in the endogenous treatments and the exogenous treatments are very similar. In particular, the use and the effectiveness of the exclusion institution are very similar. This suggests that the voting process and self-selection into the institution do not play a major role compared to the effect of the institution itself. A plausible explanation for this is that the exclusion mechanism is perceived as a relatively strong institution that is effective not only for particularly cooperative groups but, once it is implemented, for most groups.

Our results for the exogenous treatments also largely confirm the findings of previous studies (Maier-Rigaud et al., 2010; Cinyabuguma et al., 2005). The average contribution rate under the exclusion institution (73 percent in *B10-exo* and 71 percent in *B8-exo*) is slightly lower than the 80 percent found by Maier-Rigaud et al. (2010) and the 90 percent found by Cinyabuguma et al. (2005). The reason for this may be that exclusion in these studies had more severe consequences than in our setting.

[17] Regression analyses that additionally control for sample characteristics also show no significant differences between the endogenous and the exogenous treatments.

Figure 4. Comparison of contribution rates between endogenous and exogenous treatments



The figure shows average contributions over time, measured in percent of endowment, in game A (blue) and in game B (red) by treatment. The dashed lines indicate that data points are based on only few observations (N < 5). Observations in the exogenous treatment are marked by triangles. Observations in the endogenous treatment are marked by circles. The bars depict average group size for groups playing game B in the respective phase, for the endogenous treatments (empty bars) and the exogenous treatments (grey bars).

6. Discussion and conclusion

While monetary punishment has been extensively studied in the economics literature, ostracism has received much less attention and, to the best of our knowledge, the endogenous choice of an ostracism institution has not been studied previously at all. With our design, we can test if experimental groups implement an exclusion institution when they have a choice, how the choice affects cooperation and payoffs, if and how supporters and opponents of the institution differ, and how an institutional cost affects behavior. We also provide a comparison between an exclusion institution that has been chosen endogenously and one that is exogenously imposed. The behavior in the experiment certainly is noisier and more fluctuating than in theory. An important reason arguably is that the theory assumes common knowledge of preferences while the players in the

experiment have at least incomplete knowledge. Thus they need to make inferences about the co-players' preferences over the course of play and deal with the remaining uncertainty. Nevertheless, the behavior is far from random and shows remarkable stability with regard to voting between games, contributions, and the exclusions of players. Since our experimental design is not trivial, it is reassuring that our results confirm important findings from the previous literature. Like previous studies of ostracism (Masclet, 2003; Cinyabuguma et al., 2005; Maier-Rigaud et al., 2010), we find, in all of our experimental conditions, that the exclusion institution increases contributions to the public good. Subjects who were excluded or who received a vote for exclusion adjusted their contributions closer to the group average in later rounds (Masclet, 2003; Cinyabuguma et al., 2005).

The novel feature of our experiment is the endogenous choice of the exclusion institution, both when the institution is costless and when there is a cost. We show that the players' institutional choice can be better explained by assuming social preferences than by the standard model of purely selfish players. The behavior in the experiment closely resembles the predictions of the social preferences models for heterogeneous groups with some, but not only, social players. If the number of social players is high enough, they implement and use the exclusion institution to exclude the selfish players from the group and cooperate thereafter. The experimental results show that the subjects who vote for the exclusion institution contribute significantly more than those who vote against it, but only when the exclusion institution is actually implemented. If the number of social players is too low to implement the exclusion institution, the contributions of the supporters and the opponents of the institution are similarly small, just as the social preferences models predict. Two factors reduce the chances for cooperation in this case: first, the share of social players within the group is smaller and, second, the social players do not have the exclusion institution available to exclude the other players from the group. Once implemented, the exclusion institution is exclusively used to exclude the lowest contributors, which is also in line with the theoretical predictions. The support for the exclusion institution is lower when there is an institutional cost, but a significant number of players still vote in favor of it. The support becomes stronger over time, especially after the first phase, when players accustom themselves with their group and the game becomes closer to the one in which common knowledge is assumed. The increasing support for the exclusion institution is inconsistent with the predictions for groups consisting of purely selfish players only or social players only. It is, however, consistent with predictions for groups in which some players are social and some are not.

The results help to improve our understanding of the formation of institutions, the role of social preferences in this process, and how an institutional cost affects the institutional choice. Obviously, the institution formation process in the real world is not as clear-cut as in the experiment and typically the circumstances of a particular setting determine whether exclusion of group members is possible or not. The simplification of the process, however, allows us to compare groups that choose differently and individuals who vote differently. The results can help to explain why ostracism is widely used in virtually all societies around the world. With this, our study contributes to the growing literature suggesting that human preferences are heterogeneous and have a significant influence, not only on individual behavior under specific circumstances, but also on how collectives build their institutions to regulate social life.

The comparison of the endogenous treatments with the exogenous control treatments shows that the effects of the exclusion institution on cooperation, once it is implemented, are very similar. This suggests that the effect of the institution itself is more important than the sorting and signaling that comes with endogenous choice. The relative importance of these different effects is likely to depend on the interplay between the strength of the institution and the voting rule. For example, requiring a qualified majority or unanimity rather than simple majority may sort groups differently and send a different signal to the members. This might be a fruitful area for future research.

Acknowledgements. The work was financially supported by the European Union (EU) Horizon 2020 program, action ERC-2014-STG, Project HUCO, grant number 636746.

References

Akpalu, Wisdom and Peter Martinsson (2011), Ostracism and Common Pool Resource Management in a Developing Country: Young Fishers in the Laboratory, *Journal of African Economies* 21(2), 266–306.

Baland, Jean-Marie, Lata Gangadharan, Pushkar Maitra and Rohini Somanathan (2017), Repayment and Exclusion in a Microfinance Experiment, *Journal of Economic Behavior and Organization* 137, 176-190

Barrett, Scott and Astrid Dannenberg (2017), Tipping Versus Cooperating to Supply a Public Good, *Journal of the European Economic Association* 15(4), 910–941.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann (2011), A within-subject analysis of other-regarding preferences, *Games and Economic Behavior* 72(2), 321-338.

Boehm, Christopher (1986), Capital Punishment in Tribal Montenegro: Implications for Law, Biology, and Theory of Social Control, *Ethology and Sociobiology* 7(3-4), 305-320.

Bolleyer, Nicole and Anika Gauja (2015), Legal Conceptions of Organizational Membership: Implications for Intra-Party Dynamics and Democracy, Paper prepared for the Political Studies Association (UK) Annual Conference 2015.

Bowles, Samuel and Herbert Gintis (2003), Origins of human cooperation. In P. Hammerstein (Ed.), Dahlem workshop report. Genetic and cultural evolution of cooperation (pp. 429-444). Cambridge, MA, US: MIT Press.

Brekke, Kjell Arne, Karen Evelyn Hauge, Jo Thori Lind, Karine Nyborg (2011), Playing with the good guys. A public good game with endogenous group formation, *Journal of Public Economics* 95, 1111-1118.

Charness, Gary and Chun-Lei Yang (2014), Starting small toward voluntary formation of efficient largegroups in public goods provision, *Journal of Economic Behavior & Organization* 102, 119–132.

Chaudhuri, Ananish (2011), Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature, *Experimental Economics* 14(1), 47-83.

Cinyabuguma, Matthias, Talbot Page and Louis Putterman (2005), Cooperation under the threat of expulsion in a public goods experiment, *Journal of Public Economics* 89(8), 1421–1435.

Croson, Rachel, Enrique Fatas, Tibor Neugebauer, Antonio J. Morales (2015), Excludability: A laboratory study on forced ranking in team production, *Journal of Economic Behavior & Organization* 114, 13–26.

Dal Bó, Pedro (2014). Experimental Evidence on the Workings of Democratic Institutions. In S. Galiani and I. Sened (Eds.), Institutions, Property Rights, and Economics Growth: the Legacy of Douglass North, pp. 266-288. Cambridge University Press: Cambridge.

Dal Bó, Ernesto, Pedro Dal Bó, and Erik Eyster (2018), The Demand for Bad Policy When Voters Underappreciate Equilibrium Effects, *Review of Economic Studies* 85(2), 964-998.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman (2010), Institutions and Behavior: Experimental Evidence on the Effects of Democracy, *American Economic Review* 100(5), 2205–2229.

Dannenberg, Astrid and Carlo, Gallier (2019), On the Endogenous Choice of Institutions: A Survey of Experimental Research, Working Paper.

Davis, Brent J. and David B. Johnson (2015), Water Cooler Ostracism: Social Exclusion as a Punishment Mechanism, *Eastern Economic Journal* 41, 126–151.

Ehrhart, K.-M., Keser, C., (1999), Mobility and Cooperation: On the Run, mimeo.

Ertan, Arhan, Talbot Page, and Louis Putterman (2009), Who to punish? Individual decisions and majority rule in mitigating the free rider problem, *European Economic Review* 53(5), 495–511.

Fehr, Ernst and Simon Gächter (2000), Cooperation and Punishment in Public Goods Experiments, *American Economic Review* 90(4), 980-94.

Fehr, Ernst and Klaus M. Schmidt (1999), A Theory of Fairness, Competition and Cooperation, *Quarterly Journal of Economics* 114(3), 817-868.

Fehr, Ernst and Tony Williams (2017), Creating an Efficient Culture of Cooperation, Working Paper No. 267, University of Zurich, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3062528.

Feinberg, Matthew, Robb Willer, and Michael Schultz (2014), Gossip and Ostracism Promote Cooperation in Groups, *Psychological Science* 25(3), 656-664.

Fischbacher, Urs (2007), Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments, *Experimental Economics* 10(2), 271-279.

Gächter, Simon, Elke Renner, and Martin Sefton (2008), The Long-Run Benefits of Punishment, *Science* 322, 1510-1511.

Gruter, Margaret (1986), Ostracism on Trial: The Limits of Individual Rights, *Ethology and Sociobiology* 7(3-4), 271-279.

Gruter, Margaret and Roger D. Masters (1986), Ostracism as a Social and Biological Phenomenon: An Introduction, *Ethology and Sociobiology* 7(3-4), 149-158.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach (2006), The Competitive Advantage of Sanctioning Institutions, *Science* 312, 108-111.

Kopányi-Peuker, Anita, Theo Offerman, and Randolph Sloof (2018), Team production benefits from a permanent fear of exclusion, *European Economic Review* 103, 125–149.

Maier-Rigaud, Frank P., Peter Martinsson, and Gianandrea Staffiero (2010), Ostracism and the provision of a public good: experimental evidence, *Journal of Economic Behavior & Organization* 73(3), 387–395.

Markussen, Thomas, Louis Putterman, and Jean-Robert Tyran (2014), Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes, *Review of Economic Studies* 81(1), 301–324.

Masclet, David (2003), Ostracism in work teams: a public good experiment, *International Journal of Manpower*, 24(7), 867-887.

Nyborg, Karine (2017), Reciprocal climate negotiators, *Journal of Environmental Economics and Management*, doi.org/10.1016/j.jeem.2017.08.008.

Ostrom, Elinor (1990), Governing the commons. The evolution of institutions for collective action, Cambridge Univ. Press, Cambridge.

Ostrom, Elinor, James Walker, and Roy Gardner (1992), Covenants With and Without a Sword: Self-Governance is Possible, *American Political Science Review* 86(2), 404-17.

Page, Talbot, Louis Putterman, and Bulent Unel (2005), Voluntary association in public goods experiments: reciprocity, mimicry, and efficiency, *Economic Journal* 115, 1032–1053.

Putnam, Robert D., Robert Leonardi, Raffaella Y. Nanetti (1994), Making Democracy Work: Civic Traditions in Modern Italy. Princeton University Press.

Rabin, Matthew (1993), Incorporating fairness into game theory and economics, *American Economic Review* 83(5), 1281–1302.

Solda, Alice and Marie Claire Villeval (2018), Exclusion and reintegration in a social dilemma, *Economic Inquiry*, doi:10.1111/ecin.12720.

Strahilevitz, Lior Jacob (2003), Charismatic Code, Social Norms, and the Emergence of Cooperation on the File-Swapping Networks, *Virginia Law Review* 89, 505–595.

Sutter, Matthias, Stefan Haigner, and Martin G. Kocher (2010), Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations, *Review of Economic Studies* 77(4), 1540–1566.

Traxler, F./Blaschke, S./Kittel, B. (2001): National Labour Relations in Internationalized Markets. A Comparative Study of Institutions, Change and Performance, Oxford.

Tyran, Jean-Robert and Lars P. Feld (2002), Why People Obey the Law: Experimental Evidence from the Provision of Public Goods. CESifo Working Paper Series No. 651; U of St. Gallen, Econ. Discussion Paper No. 2001-14.

Tyran, Jean-Robert and Lars P. Feld (2006), Achieving Compliance when Legal Sanctions are Non-deterrent, *Scandinavian Journal of Economics* 108 (1), 135–156.

Zippelius, Reinhold (1986), Exclusion and Shunning as Legal and Social Sanctions, *Ethology and Sociobiology* 7(3-4), 159-166.

Supplementary material

for

Voting on the Threat of Exclusion in a Public Goods Experiment

Astrid Dannenberg[a]*, Corina Haita-Falah[a], and Sonja Zitzelsberger[a]

[a] Department of Economics, University of Kassel, 34117 Kassel, Germany. Emails: dannenberg@uni-kassel.de, corina.haita@gmail.com, sonja.zitzelsberger@uni-kassel.de

* Corresponding author

Appendix A: Theoretical predictions – This part contains the derivation of the theoretical predictions based on standard preferences and two social preference models: the inequality aversion model by Fehr and Schmidt (1999) and the reciprocity model by Rabin (1993).

Appendix B: Supplementary data analysis – This part presents tables and figures that support additional results discussed in the paper.

Appendix C: Experimental Instructions – This part reproduces the English translations of the original experimental instructions presented to the subjects in the German language. For the sake of space, we only present the instructions for the treatment *B8*. The instructions for the other treatments are very similar.

A. Theoretical predictions

We consider a linear public goods game with *n* players, in which the payoff of player *i* is given by:

$$\pi_i = E_p - g_i + a\sum_{j=1}^{n} g_j, \qquad i = 1,...,n \tag{A.1}$$

where E_p denotes the endowment of the players if they play game $p \in \{A,\ B10,\ B8\}$,[1] g_i is the contribution of player *i* and $0 < a < 1$ is the marginal per capita return (MPCR) from the public good.

While our experiment consists of four identical phases in which players vote for game A or game B, we analyze next only the equilibria of one phase. This means that given the choice of a game, we analyze the equilibria of that game over the five rounds of play that constitute one phase. Given the equilibria of each game, we then establish the prediction for the preference between the two games that govern players voting behavior in one phase.

A.1 Standard preferences

With standard preferences, each player maximizes the payoff function in (A.1). Because $a < 1,$ the dominant strategy of each player in the stage-game is to contribute zero. Therefore, the Nash equilibrium of the stage-game is $g_i = 0, \forall i = 1,...,n$. The subgame perfect Nash equilibrium (SPNE) can be derived by backward induction, starting from the last round of play. In round *T* the unique Nash equilibrium is to contribute zero, regardless of the history of play. At $T - 1$ every player anticipates that at *T* everyone will contribute zero. Given this, at $T - 1$ everyone contributes zero. This argument continues until the first round. It follows that in each round of the game the unique Nash equilibrium of the stage game is played. This is true for all three games in our experiment. Therefore, contributing zero in each round is the unique SPNE of the game and in each round, the SPNE payoff of each player is E_p. Thus, a player is indifferent between A and B10 and prefers A to B8 because the former gives a larger payoff. Therefore, under standard preferences game B8 is never played, while game B10 is played with 50% probability, due to the indifference result. If game B10 is chosen, then exclusion can be part of the equilibrium since exclusion in our setting is costless and, thus, players are indifferent between excluding and not excluding a player from the group. Therefore, any configuration of votes and group sizes can be part of an equilibrium.

[1] In our experiment we have either $E_B = E_A$ or $E_B = E_A - 2$, depending on the treatment condition.

A.2 Inequality-averse preferences (Fehr and Schmidt, 1999)

Inequality-averse preference of Fehr & Schmidt (1999) assumes the following utility function:

$$U_i(\pi_i) = \pi_i - \alpha_i \frac{1}{n-1}\sum_{j\neq i}\max\{(\pi_j - \pi_i),0\} - \beta_i \frac{1}{n-1}\sum_{j\neq i}\max\{(\pi_i - \pi_j),0\}, \qquad \text{(A.2.1)}$$

where the material payoff π_i is given by (A.1), if player *i* is part of a group of $n \in \{2,3,4,5\}$ players who play the public good game, or $\pi_i = E_p$, if the player is excluded from the public good. The common knowledge parameters $\alpha_i \geq \beta_i$ with $0 \leq \beta_i < 1$ measure the aversion to inequality: α_i, aversion to disadvantageous inequality and β_i, aversion to advantageous inequality.

For the ease of exposition, let us analyze each game separately.

Game A

Case 1: If there exists at least one group member with low enough advantageous inequality aversion, i.e. $\beta_i < 1 - a = 0.6$, referred to as selfish player, then all members choose $g_i = 0$. This is the unique equilibrium and the corresponding utilities are

$$U_i = E_A, \forall i \qquad \text{(A.2.2)}$$

To see that this is the unique equilibrium, we show that it is a dominant strategy for the selfish player to contribute zero. Let the total contribution of the other *n-1* players be $G_{-i} \geq 0$. Then, the utility of player *i* from contributing g_i, given that the contribution of each of the other *n-1* players is larger than g_i, is:

$$U(g_i) = E_A - g_i + a(G_{-i} + g_i) - \beta_i \frac{1}{n-1}\left(G_{-i} - (n-1)g_i\right) = E_A - (1 - a - \beta_i)g_i + \left(a - \beta_i \frac{1}{n-1}\right)G_{-i} \qquad \text{(A.2.3)}$$

Since $1 - a - \beta_i > 0$ for $\beta_i < 0.6$, it is obvious that $U(g_i)$ is maximized for $g_i = 0$ regardless of G_{-i}. With our parameter choice, due to the disadvantageous inequality aversion, the other members will also contribute zero, according to Proposition 4b in Fehr & Schmidt (1999) (for one selfish player and a=0.4, the condition from the proposition is fulfilled). Hence, zero contribution is the unique SPNE of the repeated game. The utility level any player *i* over the entire five rounds is given by:

$$U_i = 5E_A, \quad \forall i = 1,...,5 \qquad \text{(A.2.4)}$$

Case 2: If all members are sufficiently averse to advantageous inequality (conditional cooperators), i.e. $\beta_i \geq 1 - a = 0.6$, then all members choose the same contribution $g_i = g \in [0, E_A]$. It can also be shown that no member has an incentive to unilaterally reduce her contribution. Assume member *k* deviates by contributing $g - 1$. Then,

$$U_k(g-1) = E_A - (g-1) + a\big((n-1)g + g - 1\big) - \frac{1}{n-1}\beta_k(n-1) = \\ E_A + (na-1)g + (1-a-\beta_k) \le E_A + (na-1)g = U_k(g), \tag{A.2.5}$$

for any $\beta_k \geq 0.6$ and $a = 0.4$. So, deviation is weakly unprofitable. In this case, the individual equilibrium utility level over all 5 rounds, if players coordinate in each round on the same equilibrium, is:

$$U_i = 5 \times [E_A + (na-1)g], \quad \forall i = 1,...,5 \tag{A.2.6}$$

In summary, in game A, all players contribute zero if there is a least one selfish player, or may contribute the same positive amount if all players are sufficiently averse to advantageous inequality.

Game B

Case 1: For simplicity, let us first assume that there is only one selfish player with $\beta_i < 0.6$ and $n-1$ conditional cooperators with $\beta_j \geq 0.6, j \neq i$.[2] We start analyzing the game from the last round (the fifth). Because in this round there is no threat of exclusion, it is the dominant strategy for player *i* to contribute zero (see equation (A.2.3)). The question, then, arises whether it pays off for the conditional cooperators to exclude player *i* before the fifth round, anticipating her defection. For this, we compare the resulting payoffs assuming that all players contribute *g* in the first four rounds. The utility of a conditional cooperator if the selfish player is not excluded before the 5th round is:

U_j(NOT exclude before the 5th round) =

$$\underbrace{\qquad\qquad}_{\text{all players contribute } g \text{ in the first 4 rounds}} \quad \underbrace{E_B}_{\text{... players contribute 0 in the last round}} = 5E_B + 4(na-1)g \tag{A.2.7}$$

and the utility if the selfish player is excluded before the fifth round is:

U_j(exclude before round 5) = $\underbrace{\qquad\qquad}_{\text{5 players in rounds 1 to 4}}$ $\underbrace{\qquad\qquad}_{\text{player } i \text{ excluded in the 5th round}}$

$$5E_B + 4(na-1)g + g((n-1)a-1)\left(1 - \frac{1}{n-1}\beta_j\right) \tag{A.2.8}$$

Hence, it pays off for the conditional cooperator to vote for the exclusion of the selfish player before the fifth round ((A.2.8) is larger than (A.2.7) for any $\beta_j < 1$). Anticipating the exclusion, she

[2] Note that if there were two selfish players, they would both defect (see the condition from Proposition 4b in Fehr & Schmidt (1999) which is fulfilled for our parameter values). Thus, the analysis would become complicated at the exclusion decision because both players cannot be excluded at once by the remaining three players. Therefore, we resort to the simplified analysis with one selfish player to give the flavor of the choice between the two games by players with different social preferences.

may have incentive to defect already in round 4. Indeed, if player *i* contributes *g* in rounds 1-3, contributes $0 \le g_i < g$ in round 4 and is excluded from round 5, then her utility is:

U_i(contribute rounds 1-3, defect in round 4) =

$$3\times\left(E_B+(na-1)g\right)+\underbrace{E_B-(1-a)g_i+(n-1)ag-(n-1)\frac{1}{n-1}\beta_i(g-g_i)}_{\text{round 4 defection}}+\underbrace{E_B-(n-1)\frac{1}{n-1}\alpha_i((n-1)a-1)g}_{\text{round 5 excluded}}=$$

$$5E_B+((4n-1)a-3-\beta_i-((n-1)a-1)\alpha_i)g+(\beta_i-1+a)g_i \tag{A.2.9}$$

If, instead, she also contributes in round 4, but is excluded in round 5, her utility reads:

$$U_i(\text{contribute rounds 1- 4})=4\times\left(E_B+(na-1)g\right)+\underbrace{E_B-(n-1)\frac{1}{n-1}\alpha_i(4(n-1)-1)g}_{\text{round 5 excluded}}=$$

$$5E_B+[4(na-1)-\alpha_i((n-1)a-1)]g \tag{A.2.10}$$

Comparing (A.2.9) to (A.2.10), it can be shown that, since $\beta_i < 0.6$, defection in round 4 pays off for any $0 \le g_i < g$. Moreover, since the defection payoff given by (A.2.9) is strictly decreasing in g_i for any $\beta_i < 0.6$, defection in round 4 means $g_i = 0$.

Again, the common knowledge assumption allows the conditional cooperators *j* to anticipate the behavior of player *i* and exclude her before round 4. We verify next if such an exclusion threat is credible. The utility of player *j* if player *i* is excluded before round 4 is:

U_j(exclude *i* before round 4) =

$$3\times\left(E_B+(na-1)g\right)+2\times\left(E_B+((n-1)a-1)g-\frac{1}{n-1}\beta_j((n-1)a-1)g\right) \tag{A.2.11}$$

The utility of player *j* if player *i* is not excluded before round 4 and player *i* contributes zero in round 4 (as shown above) and is excluded only in round 5, is:

U_j(NOT exclude *i* before round 4) =

$$3\times\left(E_B+(na-1)g\right)+\underbrace{E_B+((n-1)a-1)g-\frac{1}{n-1}\alpha_j g}_{\text{round 4}}+\underbrace{E_B+((n-1)a-1)g-\frac{1}{n-1}\beta_j((n-1)a-1)g}_{\text{round 5}}$$

(A.2.12)

Exclusion is credible if (A.2.11) is larger than (A.2.12). This is equivalent to $\beta_j((n-1)a-1) < \alpha_j$, which is true for our MPCR $a = 0.4$, the group size of $n = 5$ and $\beta_j \le \alpha_j$.

Hence, the threat of exclusion before round 4 is credible and since it pays off for player *i* to defect in round 4, she will be excluded right before this round. By backward induction reasoning, the equilibrium of game B with one selfish player is that the selfish player is excluded at the first opportunity, i.e. after the first contribution round, and the conditional cooperators contribute an

amount $g \in [0, E_B]$ in every round until the end of the game (see the discussion from game A, Case 2). Clearly, the selfish player will defect in the first round and, consequently, all the other players will contribute zero (see Proposition 4 in Fehr & Schmidt (1999)). Then, the equilibrium utilities over the five rounds of play are as follows. The utility of the selfish player is

$$U_i = 5E_B - 4((n-1)a-1)\alpha_i g \tag{A.2.13}$$

and the utility of the conditional cooperator is:

$$U_j = 5E_B + 4((n-1)a-1)\left(1-\frac{\beta_j}{n-1}\right)g \tag{A.2.14}$$

Case 2: All players in the group have $\beta_i \geq 0.6$. Hence, similarly as in game A, all players contribute the same weakly positive amount $g_i = g \in [0, E_B]$.[3] Note that, unlike in the case with one selfish player, the absence of the possibility for exclusion after round 5 does not create incentive for deviation. Instead, in round 5 everyone still contributes $g \in [0, E_p]$. Hence, the equilibrium payoffs in game B when all players have $\beta_i \geq 0.6$ is:

$$U_i(g) = 5 \times \left(E_B + (na-1)g\right). \tag{A.2.15}$$

Note that the equilibrium of this game entails coordination on a certain contribution level. A natural focal point for this is $g = E_B$.[4]

Given the equilibria derived above, we can now derive prediction regarding the choice between game A and game B, when players are inequality averse.

Table A.2.1 summarizes the equilibrium utility levels for the two cases discussed above.

Table A.2.1 Equilibrium utilities

	Game A	**Game B**
Case 1: $\exists \beta_i < 0.6$	$U_i^A = 5E_A, \forall i$	$U_i^B = 5E_B - 4((n-1)a-1)\alpha_i g$

[3] Furthermore, it can be shown that some members, depending on their social preference parameters, may be tolerant to deviations below g. In particular, a cooperator may not vote out a player who contributes less than g, but at least $g[\alpha_j - \beta_j((n-1)a-1)]/[\alpha_j + (n-1)a]$. However, no player with $\beta_i > 0.6$ has an incentive to contribute less than g.

[4] There exists experimental evidence that points to the fact that coordination is harder in larger than in smaller groups. Therefore, one may conceive formation of smaller, more coordinated groups as other reasons for exclusion. Indeed, one can show that, for example, if the group of 5 players coordinate on a lower contribution than a fraction of the contribution expected in a group of 4 players, i.e., $g_5 < g_4(4a-1)(n-1.6)/[(5a-1)(n-1)]$, then one player is voted out by another player if the latter's aversion to advantageous inequality is low enough. Note that this analysis does not say anything about who will be voted. Consequently, voting on exclusion for the sole purpose of reducing the group size in order to increase the chance of coordination on higher contributions is in itself a matter of coordination.

		$U_j^B = 5E_B + 4((n-1)a-1)\left(1-\frac{\beta_j}{n-1}\right)g$
Case 2: $\beta_j \geq 0.6, \forall j$	$U_j^A = 5\times[E_A + (na-1)g],\ \forall j$	$U_j^B = 5\times[E_B + (na-1)g],\ \forall j$

Case 1: In this case we have to consider separately the choice of the player with $\beta_i < 0.6$ and the choice of the players with $\beta_j \geq 0.6$. For this, we have to compare U_i^A with U_i^B and U_j^A with U_j^B. In particular, player *i* prefers game A if and only if $5(E_A - E_B) > -4((n-1)a-1)\alpha_i g$, which is true since $E_A \geq E_B$. Therefore, a selfish player strictly prefers game A in which the exclusion institution is not available. In turn, player *j* prefers game A if and only if $5(E_A - E_B) > 4((n-1)a-1)\left(1-\frac{\beta_j}{n-1}\right)g$, with $\beta_j \geq 0.6$. Here we have to consider separately the two exclusion games. If game B10 is the alternative ($E_A = E_B$) and $g > 0$, then game B10 is strictly preferred if $\beta_j < n-1 = 4$, which is true for all $\beta_j < 1$. This means that when there is no cost of the exclusion institution, a conditional cooperator always votes for it. If game B8 is the alternative ($E_A > E_B$) and $g > 0$, then it is strictly preferred to A if $\beta_j < (n-1) - \frac{5(E_A - E_B)(n-1)}{4((n-1)a-1)g} \equiv \overline{\beta_j}$. Thus, a conditional cooperator chooses game B8 if he is not too averse to advantageous inequality. However, given that $0.6 \leq \beta_j < 1$, for this condition to be relevant it is necessary that $g \geq \frac{5(E_A - E_B)(n-1)}{4(n-1.6)((n-1)a-1)} \approx 4.9$. Hence, for our parameter values, B8 is preferred if $g \geq 5$ and β_j is sufficiently low. Note that for $g \geq 6$ we have that $\overline{\beta_j} > 1$ and, thus, the condition for B8 to be preferred is always fulfilled, i.e. the advantageous inequality aversion no longer plays a role. Finally and for completion, if the conditional cooperators coordinate on the inefficient equilibrium $g = 0$, then all players are indifferent between A and B10, but prefer A to B8.

To summarize, a selfish player always chooses game A, while a conditional cooperator will choose A only if he is very averse to advantageous inequality. However, the aversion to advantageous inequality will only play a role, for our choice of parameters, if the contribution of the conditional cooperators is sufficiently low. In all other cases a conditional cooperator prefers game B.

Case 2: If $\beta_i \geq 0.6, \forall i$, then game A is preferred if $E_A > E_B$ and players are indifferent between A and B if $E_A = E_B$. Hence, A is preferred to B8 and players are indifferent between A and B10. Note

that this conclusion assumes that players coordinate on the same contributions in game A as in game B10 and the contributions in A are at least as large as the contributions in B8.[5]

A.3 Reciprocity preferences (Rabin, 1993)

Rabin (1993) assumes that the utility of player *i* is given by:

$$u_i = \pi_i + \beta_i R_i,$$

where π_i is the material payoff given by (A.1). Parameter β_i is the weight attributed to reciprocation and R_i is the reciprocation concern term. Furthermore, the model assumes common knowledge of preferences. We follow the extension of the Rabin (1993) model to an *n*-player public goods game developed by Nyborg (2017).

Case 1: Symmetric players. If all *n* symmetric players have reciprocal concerns, i.e. $\beta_i = \beta > 0, \forall i = 1,...,n$ and because they are identical with respect to their reciprocity preferences, then there is no difference between game A and game B. This is the case because in game B there is no reason for a player to be excluded by her group given the symmetric equilibria that we derive below. Thus, games A and B have the same equilibria.

The maximum payoff that *i* can secure for $j \neq i$, based on i's beliefs about j's contribution, is given by

$$\pi_{ij}^{max} = E_p - g_j + a(G_{-i} + E_p) \tag{A.3.1}$$

and the minimum payoff that *i* can secure for *j* is

$$\pi_{ij}^{min} = E_p - g_j + a(G_{-i} + 0), \tag{A.3.2}$$

where $G_{-i} = \sum_{\substack{l=1, \\ l \neq i}}^{n} g_l$. Then, the equitable payoff is defined as:

$$\pi_{ij}^{e} = \frac{1}{2}\left(\pi_{ij}^{max} + \pi_{ij}^{min}\right) = E_p - g_j + aG_{-i} + \frac{aE_p}{2}. \tag{A.3.3}$$

Using (A.3.1), (A.3.2) and (A.3.3), we can calculate the kindness of player *i* towards player *j*:

$$f_{ij} = \frac{\pi_j(g_i, G_{-i}) - \pi_{ij}^{e}}{\pi_{ij}^{max} - \pi_{ij}^{min}} = \frac{E_p - g_j + a(G_{-i} + g_i) - (E_p - g_j + aG_{-i} + \frac{aE_p}{2})}{aE_p} = \frac{g_i}{E_p} - \frac{1}{2}, \quad i \neq j, \tag{A.3.4}$$

[5] This is because the full cooperation contributions are larger in game A than in B8 given the larger endowment in A compared to B8.

where $\pi_j\left(g_i, G_{-i}\right)$ is the material payoff of player *j* as a function of i's contribution g_i and given i's beliefs about others' contributions, G_{-i}. Symmetrically, the beliefs of player *i* about *j's* kindness towards *i* write:

$$\tilde{f}_{ji} = \frac{g_j}{E_p} - \frac{1}{2},\ i \neq j.$$

From equation (A.3.4) we can see that player *i* is neither kind nor unkind with player *j* if player *i* secures the equitable payoff for player *j,* i.e. $f_{ij} = 0$. If $f_{ij} < 0$, then player *i* is "unkind" as she is securing for *j* less than her equitable payoff. If $f_{ij} > 0$, then player *i* is "kind" by securing for *j* more than her equitable payoff.

Using equation (A.3.4) and $f_{ij} = f_{il},\ \forall j, l \neq i$, we can write the reciprocal term in the utility function of player *i* as defined in Nyborg (2017):

$$R_i = \frac{1}{n-1}\left(\sum_{j\neq i}\tilde{f}_{ji} + \sum_{j\neq i} f_{ij}\,\tilde{f}_{ji}\right) = \frac{1}{n-1}\left[\sum_{j\neq i}\left(\frac{g_j}{E_p} - \frac{1}{2}\right) + \sum_{j\neq i}\left(\frac{g_i}{E_p} - \frac{1}{2}\right)\left(\frac{g_j}{E_p} - \frac{1}{2}\right)\right] = \left(\frac{1}{E_p}\frac{G_{-i}}{n-1} - \frac{1}{2}\right)\left(\frac{g_i}{E_p} + \frac{1}{2}\right), \tag{A.3.5}$$

where the sums are over $j = 1,...,i-1,i+1,...,n.$ The reciprocal utility function of player *i* as a function of own and others' contributions is:

$$u_i(g_i, G_{-i}) = E_p - g_i + a(G_{-i} + g_i) + \beta\left(\frac{1}{E_p}\frac{G_{-i}}{n-1} - \frac{1}{2}\right)\left(\frac{g_i}{E_p} + \frac{1}{2}\right) \tag{A.3.6}$$

The first order condition writes:

$$\frac{\partial u_i(g_i, G_{-i})}{\partial g_i} = -1 + a + \frac{\beta}{E_p}\left(\frac{1}{E_p}\frac{G_{-i}}{n-1} - \frac{1}{2}\right) = 0,$$

which gives the following reaction function:

$$g_i = \begin{cases} E_p, & \text{if } \dfrac{1}{E_p}\dfrac{G_{-i}}{n-1} > \dfrac{1}{2} + \dfrac{E_p}{\beta}(1-a) \\ 0, & \text{if } \dfrac{1}{E_p}\dfrac{G_{-i}}{n-1} < \dfrac{1}{2} + \dfrac{E_p}{\beta}(1-a) \end{cases} \tag{A.3.7}$$

For $\dfrac{1}{E_p}\dfrac{G_{-i}}{n-1} = \dfrac{1}{2} + \dfrac{E_p(1-a)}{\beta}$ the player is indifferent between contributing anything between zero and E_p. Player *i* contributes her full endowment if the average contribution of the other players as a share

of the individual endowment is strictly larger than half, and contributes nothing if the other players' average contribution relative to the initial endowment is well below one half.

Using equation (A.3.6), simple algebra shows that $u_i(0,0) > u_i(g_i,0), \forall g_i > 0$ if $\beta > 2E_p(-1+a)$, which holds for $\beta > 0$ and $a < 1$. This means that for any player *i*, $g_i = 0$ is the best response to $G_{-i} = 0$. This makes zero contribution an equilibrium. Similarly it can be shown that that full contribution is a best response to the full contribution of the remaining reciprocal players if $\beta > 2E_p(1 - a)$. This is obtained by comparing $u_i(E_p,(n-1)E_p)$ with $u_i(g_i,(n-1)E_p), \forall g_i < E_p$.

Thus, we have the following two pure-strategy Nash equilibria in both games:[6]

$(i)\ g_i = 0, \forall i = 1,...,n$

$(ii)\ g_i = E_p, \forall i = 1,...,n$ if $\beta > 2E_p(1-a)$

Hence, for $\beta > 2E_p(1 - a)$, the stage-game is a coordination game with two Pareto ranked equilibria.[7] For $\beta \leq 2E_p(1-a)$, the reciprocity game has the same unique Nash equilibrium as the game with standard preferences, i.e. the zero contribution equilibrium and it is also the SPNE of the repeated game. Therefore, exclusion can be part of the equilibrium in games B because players are indifferent between exclusion and non-exclusion. However, due to non-unique Nash equilibrium of the stage game for $\beta > 2E_p(1-a)$, we have multiple SPNE. Two of these equilibria are the repetition of each of the two stage-game equilibria.

Case 2: Asymmetric players

Let us assume for simplicity that $n-1$ players have reciprocation concern (the reciprocal players) and one player does not have reciprocation concern (the non-reciprocal player).[8] Let us further index the non-reciprocal player with *k*. Then $\beta_i = \beta > 0,\ i \neq k$ and $\beta_k = 0$. We start the analysis with game B.

---

[6] For an *n*-player prisoner's dilemma game, Nyborg (2017) shows that there is also a mixed-strategy Nash equilibrium: if the reciprocal players are sufficiently reciprocal, then they mix between defection and cooperation, while the non-reciprocal player plays defection with probability one.

[7] One may still ask whether a reciprocal player contributing less than the full contribution off the equilibrium path would be tolerated by the other players, when game B is played. As it turns out, this would require that the contribution of the deviating player is above a certain threshold and the reciprocity parameter β of the cooperating players is low enough. However, this is not consisted with the value of β for which the full contribution equilibrium exists. Therefore, a player who contributes less that the full endowment when everyone else contributes the full endowment is voted out. Nevertheless, since full contribution is the best response to full contribution, such a situation does not occur.

[8] It is also possible that there are more than one non-reciprocal players in the group. Indeed, Nyborg (2017) considers this case and finds the same Nash equilibria as we do, for the game without the exclusion institution. However, as the number of the non-reciprocal players increases, the threshold for which the reciprocal players contribute the full endowment (see below) also increases, meaning that cooperation among the reciprocal players is harder to sustain. With multiple non-reciprocal players, the analysis of game B would complicate due to the fact that only one player can be excluded per round. We find that this complication is not worth pursuing in order to get the gist of the game incorporating reciprocal preferences.

Game B

We first consider the perspective of the reciprocal players and we focus on the last round to solve the game by backward induction. Their possible actions are: exclude the non-reciprocal player before the last round and then contribute g_i or do not exclude the reciprocal player and then contribute g_i. Let us first establish the kindness of the reciprocal player $i \neq k$ towards player $j \neq i,k$. Then π_{ij}^{max} and π_{ij}^{min} are given by equations (A.3.1) and (A.3.2), respectively and π_{ij}^{e} is given by (A.3.3). Similar calculations as in the symmetric case give the kindness of player *i* towards player *j* as

$$f_{ij} = \frac{g_i}{E_p} - \frac{1}{2}, i \neq j,$$

and the beliefs of player *i* about the kindness of player *j* towards *i*,

$$\tilde{f}_{ji} = \frac{g_j}{E_p} - \frac{1}{2}, i \neq j.$$

Let us now consider the kindness of player $i \neq k$ towards player *k.* Players *i* can affect *k*'s payoff in two ways: by excluding her before the last round or by allowing her in the game. We consider each case in turn.

If players *i* do *not exclude* player *k* from the game, she obtains utility from the material payoff:

$$u_k(g_k, G_{-k}) = E_p - g_k + a(G_{-k} + g_k).$$

Since in the last round there is no threat of exclusion, it is clear that her dominant strategy is $g_k = 0$ because $a < 1$. Moreover, player *i*'s kindness towards *k* is given by $f_{ik} = \frac{g_i}{E_p} - \frac{1}{2}$ and *i*'s belief about *k*'s kindness is given by $\tilde{f}_{ki} = \frac{g_k}{E_p} - \frac{1}{2}$. This allows us to write the reciprocal term in the utility function of player $i \neq k$ as in (A.3.5):

$$R_i = \left(\frac{1}{E_p} \frac{G_{-i}}{n-1} - \frac{1}{2} \right) \left(\frac{g_i}{E_p} + \frac{1}{2} \right), \tag{A.3.8}$$

where G_{-i} includes the contribution of player *k.* Finally, we can write the reciprocal utility function of player $i \neq k$ as a function of own and others' contributions:

$$u_i(g_i, G_{-i}) = E_p - g_i + a(G_{-i} + g_i) + \beta \left(\frac{1}{E_p} \frac{G_{-i}}{n-1} - \frac{1}{2} \right) \left(\frac{g_i}{E_p} + \frac{1}{2} \right), \quad i \in \{1, \ldots \tag{A.3.9}$$

Using the same reasoning as in the symmetric case we can show that for any $i \neq k$, $g_i = 0$ is the best response to $G_{-i} = 0$. This means that zero contribution of the reciprocal players is an equilibrium.

Similarly it can be shown that full contribution of the reciprocal players is a best response to the full contribution of the remaining reciprocal players if $\beta > 2E_p(1-a)\frac{n-1}{n-3}$. This is obtained by comparing $u_i\left(E_p,(n-2)E_p\right)$ with $u_i\left(g_i,(n-2)E_p\right), \forall g_i < E_p$.

Hence, we have two pure strategy Nash equilibria:

$(i)\ g_i = 0,\ \forall i = 1,...,n$

$(ii)\ g_k = 0$ and $g_i = E_p,\ \forall i \neq k,\ i = 1,...,n$ if $\beta > 2E_p(1-a)\frac{n-1}{n-3} \equiv \beta_{not\ excl.}$

Thus, for sufficiently high reciprocity preferences, full cooperation by the reciprocal players can be sustained as a Nash equilibrium. Moreover, the threshold for which full cooperation is sustained decreases in the number of players. This means that the more reciprocal players are in the game, the easier it is for this equilibrium to exist. The corresponding utilities are:

$(i)\ u_k = E_p$ and $u_i = E_p - \frac{\beta}{4},\ \forall i \neq k$

$(ii)\ u_k = E_p + a(n-1)E_p + \frac{1}{4}\beta$ and $u_i = aE_p(n-1) + \frac{3(n-3)}{4(n-1)}\beta,\ \forall i \neq k$

If the reciprocal players *exclude* the non-reciprocal player *k* before the last round, then she has no further action in this round of the game. Therefore, she can be neither kind nor unkind to player *i*. That means that $\tilde{f}_{ki}$ is not defined. Therefore, the attitude of *i* towards player *k* is irrelevant for *i*'s utility function and $u_k = E_p$. In this case the reciprocity term in the utility function is determined by the kindness of $i \neq k$ towards the remaining $n-2$ reciprocal players and her beliefs about the kindness of these player towards herself:

$$R_i = \frac{1}{n-2}\left(\sum_{j\neq i}\tilde{f}_{ji} + \sum_{j\neq i} f_{ij}\,\tilde{f}_{ji}\right) = \left(\frac{1}{E_p}\frac{G_{-i}}{n-2} - \frac{1}{2}\right)\left(\frac{g_i}{E_p} + \frac{1}{2}\right), \qquad \text{(A.3.10)}$$

where the sums are over $j \in \{1,...,i-1,i+1,...,n-1\},\ j \neq k$. Hence, the reciprocal utility of player $i \neq k$ is:

$$u_i(g_i, G_{-i}) = E_p - g_i + a(G_{-i} + g_i) + \beta\left(\frac{1}{E_p}\frac{G_{-i}}{n-2} - \frac{1}{2}\right)\left(\frac{g_i}{E_p} + \frac{1}{2}\right), \quad i \in \{1,\ldots \qquad \text{(A.3.11)}$$

Since the non-reciprocal player is excluded, the game is now equivalent with the symmetric case analyzed above. Therefore, we have the same two pure-strategy Nash equilibria with $n-1$ players:

$(i)\ g_i = 0, \forall i \neq k,\ i = 1,...,n-1$

$(ii)\ g_i = E_p, \forall i \neq k,\ i = 1,...,n-1$ if $\beta > 2E_p(1-a) \equiv \beta_{excl.}$

Thus, if reciprocity preference is strong enough, then full cooperation by all reciprocal players can be sustained as a Nash equilibrium. The corresponding utilities are:

$$(i)\ u_i = E_p - \frac{\beta}{4}, \forall i \neq k$$

$$(ii)\ u_i = aE_p(n-1) + \frac{3}{4}\beta,\ \forall i \neq k$$

Having solved for the equilibrium of the contribution decisions in the last round, both in the case the non-reciprocal player is excluded before this stage and in the case she is not excluded, we now turn to the decision of the reciprocal players whether to exclude the non-reciprocal player right before the last contribution round. For this we will assume coordination among the reciprocal players on one of the equilibria.

First, we note that for our parameter values $\beta_{not\ excl.} > \beta_{excl.}$, since $\frac{n-1}{n-3} > 1,\ \forall n > 3$. Thus, full cooperation by the reciprocal players when the non-reciprocal player is in the game requires a higher level of reciprocal preference than in the case when the non-reciprocal player is excluded from the game. This is intuitive since the existence of the non-reciprocal player in the game decreases the social utility of the reciprocal players. Hence, it takes a high level of reciprocal preference for this decrease of utility to be offset by the reciprocation of the other reciprocal players. We also note that $\beta_{\text{not excl.}}$ depends on the number of (reciprocal) players, while $\beta_{\text{excl.}}$ does not. This is also intuitive because, in the presence of a non-reciprocal player, it becomes relevant how many other reciprocal players are in the game such that it is worth for them to cooperate fully.

To summarize, whenever $\beta > \beta_{not\ excl.}$, full contribution by the reciprocal players is an equilibrium regardless of the presence of the non-reciprocal player. Next, if $\beta_{\text{excl.}} \leq \beta \leq \beta_{\text{not excl.}}$, then the full contribution equilibrium exists only if the non-reciprocal player is excluded. Finally, if $\beta < \beta_{\text{excl.}}$, we only have the equilibrium in which all players contribute zero. Let us analyze these cases in turn by comparing all possible combinations of equilibrium outcomes for the exclusion and the non-exclusion cases:

- $\beta \leq \beta_{excl.}$. In this case the reciprocal players are indifferent between excluding and not excluding the non-reciprocal player because they contribute zero regardless of the non-reciprocal player being in the game or not.
- $\beta_{excl.} < \beta \leq \beta_{not\ excl.}$. In this case, the exclusion decision depends on which equilibrium is played if the reciprocal players exclude the non-reciprocal one. If the zero contribution equilibrium is played, then the reciprocal players are indifferent between exclusion and non-exclusion. If they coordinate on the full contribution equilibrium after the exclusion, then they strictly prefer to

exclude the non-reciprocal player. By backward induction, we obtain that the non-reciprocal player is excluded after the first round and the reciprocal players play the full contribution equilibrium thereafter. However, in the first round the reciprocal players can only play the zero contribution equilibrium.

- $\beta > \beta_{not\ excl.}$. In this case, the exclusion decision depends on the combination of equilibria that are played when the non-reciprocal player is not excluded and when she is excluded. If the reciprocal players coordinate on the full contribution equilibrium after the exclusion of the non-reciprocal player, then they are better off excluding the non-reciprocal player regardless of the equilibrium they play if the reciprocal player is not excluded.[9] If the reciprocal players coordinate on the zero contribution equilibrium regardless of whether the reciprocal player is excluded or not, then they are indifferent between exclusion and non-exclusion. Again, the backward induction reasoning obtains that reciprocal players exclude the non-reciprocal one as early as possible, i.e. in the first round of voting right after the first contribution decision. Finally, in the implausible situation in which the reciprocal players coordinate on the full contribution equilibrium in the non-exclusion case but play the zero contribution equilibrium in the exclusion case, they prefer not to exclude the non-reciprocal player.

From the above discussion, we can conclude that, for the threat of exclusion to be credible, it must be that the reciprocal players coordinate on the full contribution equilibrium after the exclusion takes place. This, in turn, requires that the full contribution equilibrium exists, i.e. the reciprocal players are sufficiently reciprocal ($\beta > \beta_{excl.}$).

Thus, we have shown that full cooperation can be sustained (at least from the second round onwards) and the non-reciprocal player is excluded in the first voting round if the reciprocal players are reciprocal enough and they coordinate on the full contribution equilibrium.

Game A

The analysis and the outcome are identical to the one for the case of no exclusion in game B. Specifically, for each round of play we have the following Nash equilibria:

$(i)\ g_i = 0, \forall i = 1, \ldots, n$

$(ii)\ g_k = 0$ and $g_i = E_p, \forall i \neq k$ if $\beta > 2E_p(1-a)\dfrac{n-1}{n-3}$

with the corresponding utilities:

[9] In particular, the utility of a reciprocal player from full contribution equilibrium is higher if the non-reciprocal player is excluded than if she is still present in the game. To see this, compare equations (A.3.9) and (A.3.11).

$(i)\, u_k = E_p$ and $u_i = E_p - \frac{\beta}{4}, \forall i \neq k$

$(ii)\ u_k = E_p + a(n-1)E_p + \frac{1}{4}\beta$ and $u_i = aE_p(n-1) + \frac{3(n-3)}{4(n-1)}\beta,\ \forall i \neq k$

In order to conduct the comparisons between the games, we assume that players coordinate on the same equilibrium throughout the five rounds of play. Table A.3.1 shows the utilities of the reciprocal players over the 5 rounds of play for each of the two equilibria and for each game.

Table A.3.1 Equilibrium utilities of the reciprocal players over the five rounds of play

Equilibrium	Game A	Game B
$g_i = 0, \forall i \in \{1,...,n\}$ and $\forall \beta$	$5\left(E_A - \frac{\beta}{4}\right)$	$5\left(E_A - \frac{\beta}{4}\right)$
$\beta > \beta_{not\,excl.}$, $g_i = E_p$, $i \neq k$ and $g_k = 0$	$5\left(aE_A(n-1) + \frac{3(n-3)}{4(n-1)}\beta\right)$	$5aE_B(n-1) + \frac{3(5n-7)}{4(n-1)}\beta$ *
$\beta_{not\,excl.} > \beta > \beta_{excl.}$ Game A: $g_i = 0, \forall i \in \{1,...,n\}$ Game B: $g_i = E_A, \forall i \in \{1,...,n-1\}$ after round 1	$5\left(E_A - \frac{\beta}{4}\right)$	$\left(E_A - \frac{\beta}{4}\right) + 4\left(aE_B(n-1) + \frac{3}{4}\beta\right)$

*This is the result of $\underbrace{aE_B(n-1) + \frac{3(n-3)}{4(n-1)}\beta}_{\text{round 1}} + 4\underbrace{\left(aE_B(n-1) + \frac{3}{4}\beta\right)}_{\text{non-reciprocal player is excluded in rounds 2 to 5}}$

It is clear that if the reciprocal players play the zero equilibrium consistently across the two games, then they are indifferent between B10 and A, but strictly prefer A to B8. For the cases in which the two pure-strategy Nash equilibria co-exist in both games, i.e. $\beta > \beta_{not\ excl.}$ it may seem plausible to assume that the reciprocal players coordinate on the Pareto-dominant equilibrium in which they contribute their endowments. Then it can be shown that they strictly prefer B10 to A, and strictly prefer B8 to A only if the reciprocity preference is high enough, i.e. $\beta > \frac{5a(n-1)^2(E_A - E_{B8})}{6}$. Because, for our parameter values we have that $\frac{5a(n-1)^2(E_A - E_{B8})}{6} < 2E_{B8}(1-a)\frac{n-1}{n-3}$, the condition for which B8 is preferred becomes irrelevant as long as the full contribution equilibrium exists. Thus, if the full contribution equilibrium exists and it is played, then game B8 is always preferred. Next, if $\beta_{excl.} > \beta > \beta_{not\ excl.}$, then full contribution equilibrium does not exist in game A. However, the full contribution equilibrium by the reciprocal players exists in game B if the reciprocal player is excluded after the first round. This case is presented in the last row of Table A.3.1. By

comparing the payoffs, it becomes clear that both games B are preferred to game A. Finally, in the unlikely case in which the reciprocal players coordinate on full contribution in game A and on zero contribution in game B, then game A is strictly preferred for our parameter values.

It is straightforward to see that the non-reciprocal player strictly prefers game A to game B, since game A allows her to benefit from the public good while defecting in all rounds.

A.4 References

Fehr, E. & Schmidt, K. M. (1999), 'A theory of fairness, competition and cooperation', *The Quarterly Journal of Economics* pp. 817–868.

Nyborg, Karine (2017), Reciprocal climate negotiators, *Journal of Environmental Economics and Management*, doi.org/10.1016/j.jeem.2017.08.008.

Rabin, M. (1993), 'Incorporating fairness into game theory and economics', *American Economic Review* 83(5), 1281–1302

B. Supplementary data analyses

B.1 Contributions

Table B.1 shows average cooperator contributions as percentage of endowment conditional on treatment, phase, and game. Recall that "cooperators" are the non-excluded players in game B and all players but the lowest contributor(s) in game A. In all treatments and phases the average contributions of cooperators in game B are higher than in game A.

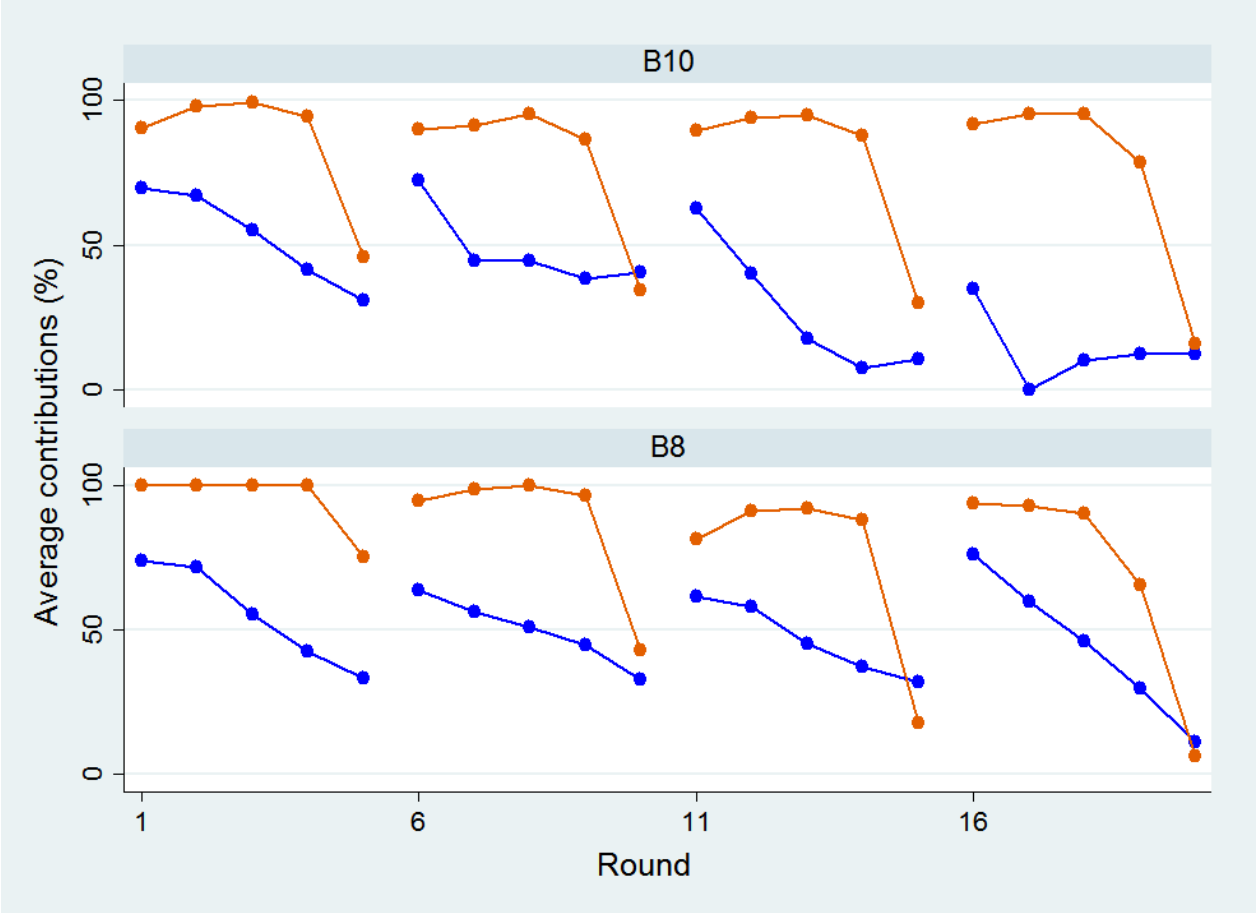
Table B.1 Average contributions of cooperators

Phase	Game	B10	B8
I	A	0.5282 (0.2804, 16)	0.5523 (0.2403, 22)
	B	0.8575** (0.0429, 7)	0.9500† (-, 1)
II	A	0.4810 (0.3601, 5)	0.4950 (0.2829, 14)
	B	0.7948* (0.1395, 18)	0.8640*** (0.0704, 9)
III	A	0.2763 (0.1465, 4)	0.4671 (0.2622, 14)
	B	0.7930† (0.1190, 19)	0.7404*** (0.1205, 9)
IV	A	0.14 (-, 1)	0.4442 (0.2265, 11)
	B	0.7536† (0.1094, 22)	0.6966*** (0.1245, 12)
Total	A	0.4777 (0.2754, 16)	0.5376 (0.2896, 16)
	B	0.7725*** (0.1134, 23)	0.7494*** (0.1618, 23)

Numbers show average cooperator contributions conditional on treatment, phase, and game. Standard deviation and number of groups are shown in parentheses. In game A the average excludes the lowest contributor(s) in the respective phase. In game B only non-excluded subjects are considered. Stars indicate significant difference in cooperator contributions between game A and game B within the same treatment and phase (MWW test statistic). Level of significance: *p < 0.10, **p < 0.05, ***p < 0.01. † indicates that statistical tests are not possible due to the low number of observations (N < 5) in at least one category.

Figure B.1 shows the average contributions of cooperators over time. It can be clearly seen that average contributions of cooperators in game B stay constant or even increase up to the third or fourth round of each phase, before they sharply drop in the last round. Average contributions of cooperators in game A decline steadily from the first round onwards.

Figure B.1 Cooperator contributions over time by treatment



The lines show average cooperator contributions, measured as percent of endowment. Contributions in game A are depicted in blue and contributions in game B are depicted in orange. In game A the average excludes the lowest contributor(s) in the respective phase. In game B only subjects that were not excluded in the respective phase are considered.

B.2 Differences between A-voters and B-voters

Table B.2 shows regression results on average contributions per phase in game B by treatment and whether or not game B is played for the first time. Having voted for game B increases average contributions in the phase when game B is played the first time by approximately 19 percentage points in *B10* and by 11 percentage points in *B8*. If game B is not played for the first time, there is no significant difference in average contributions across the phase between A-voters and B-voters.

Table B.2 Average contributions per phase in game B

	Game B is played for the first time				Game B is not played for the first time			
	(1) B10	(2) B10	(3) B8	(4) B8	(5) B10	(6) B10	(7) B8	(8) B8
Voted for game B (d)	0.1866***	0.1951***	0.1117*	0.1134*	0.0394	0.0237	0.0279	0.0179
	(0.0565)	(0.0661)	(0.0602)	(0.0559)	(0.0280)	(0.0272)	(0.0537)	(0.0448)
Average contribution (%) in previous phase		0.3864***		0.3631***		0.1231		0.2302*
		(0.0878)		(0.0918)		(0.0770)		(0.1241)
Game B in previous phase (d)						-0.0677		-0.1803**
						(0.0731)		(0.0897)
Observations	115	80	75	70	215	215	80	80

OLS estimation results (Columns (1)-(4)) and random effects GLS estimation results (Columns (5)-(8)) with standard errors in parentheses. Standard errors are clustered by group. Dependent variable is the average contribution as percentage of endowment in the phase when game B is played. Level of significance: * p < 0.10, ** p < 0.05, *** p < 0.01. Columns (2) and (4) – (8) exclude the first phase. Regressions include dummy indicators for phases. (d) indicates dummy variable.

Tables B.3 and B.4 show regression results on contribution rates when game A is played. According to Table B.3, having voted for game B has no significant effect on first round contributions when game A is played for the first time. If game A is not played for the first time, the voting preferences have no or only a small effect on first round contributions. The only statistically significant effect is found in the *B10* treatment. Here, having voted for game B decreases first round contributions by about 9 percentage points. Table B.4 shows the corresponding results for average phase contributions. If game A is played for the first time, having voted for game B does not have any significant effect on contributions. If game A is not played for the first time, having voted for game B decreases contributions by about 7 percentage points in the *B10* treatment.

Table B.3 First round contributions in game A

	Game A is played for the first time		Game A is not played for the first time			
	(1) B10	(2) B8	(3) B10	(4) B10	(5) B8	(6) B8
Voted for game B (d)	0.0203	-0.0449	-0.0267	-0.0934***	-0.0090	0.0036
	(0.0745)	(0.1169)	(0.0352)	(0.0281)	(0.0597)	(0.0437)
Average contribution (%) in previous phase				0.6616***		0.8248***
				(0.1925)		(0.0564)
Game B in previous phase (d)				-0.2840		-0.1781***
				(0.2186)		(0.0663)
Observations	80	110	50	50	195	195

OLS estimation results (Column (1)-(2)) and random effects GLS estimation results (Column (3)-(6)) with standard errors in parentheses. Standard errors are clustered by group. Dependent variable is the individual contribution as percentage of endowment in game A. Level of significance: * p < 0.10, ** p < 0.05, *** p < 0.01. Columns (3) – (6) include only phases II-IV. Regressions include dummy indicators for phases. (d) indicates dummy variable.

Table B.4 Average contributions per phase in game A

	Game A is played for the first time		Game A is not played for the first time			
	(1) B10	(2) B8	(3) B10	(4) B10	(5) B8	(6) B8
Voted for game B (d)	-0.0028	0.0982	-0.0113	-0.0716***	-0.0430	-0.0274
	(0.0547)	(0.0921)	(0.0373)	(0.0264)	(0.0412)	(0.0299)
Average contribution (%) in previous phase				0.7194***		0.7988***
				(0.1833)		(0.0456)
Game B in previous phase (d)				-0.3897***		-0.2433***
				(0.1170)		(0.0669)
Observations	80	110	50	50	195	195

OLS estimation results (Column (1)-(2)) and random effects GLS estimation results (Column (3)-(6)) with standard errors in parentheses. Standard errors are clustered by group. Dependent variable is the average contribution as percentage of endowment in the phase when game A is played. Level of significance: * p < 0.10, ** p < 0.05, *** p < 0.01. Columns (3) – (6) include only phases II-IV. Regressions include dummy indicators for phases. (d) indicates dummy variable.

B.3 Group size and exclusion of players

Table B.5 compares contributing behavior of excluded subjects and non-excluded subjects. Specifically, it shows the gap between own contribution and average contribution of the other group members for excluded and non-excluded subjects in the phase of the exclusion and the following phase. Only groups that played two subsequent games B are included in the analysis as there are only few groups that switched to game A after having played game B. The numbers indicate that previously excluded subjects adapt their contribution levels in the direction of the others' average, but they contribute still less than the others. Non-excluded subjects keep their contribution levels constant and very close to the level of the other group members.[10]

Table B.5 Individual contributions and gap between individual contribution and average contribution of other group members

Treatment	Excluded in phase t	Observations	Phase	Contributions (%)		Gap in contributions (pp)	
				Mean	Std. Dev.	Mean	Std. Dev.
B10	Yes	43	t	58	27	-26	25
			t+1	71	17	-6	17
	No	157	t	83	14	3	14
			t+1	79	13	-0.4	15
B8	Yes	11	t	47	31	-40	26
			t+1	70	15	-12	18
	No	49	t	82	14	3	10
			t+1	73	15	1	10

The table shows the average individual phase contributions as percentage of endowment and average gap between individual contribution and average contribution of other group members in percentage points (pp) for excluded and non-excluded subjects in the phase of their (non-)exclusion and the following phase for two subsequent games B (in percentage points). Phase t is one of the phases I-III, and phase t+1 one of the phases II-IV.

Table B.6 examines the likelihood of previously excluded individuals to be excluded again, when game B is played both in the previous and the current phase. Having been excluded in any of the

[10] The results also hold when we conduct the same analysis only for subjects who contributed less than the average contribution of the others. Excluded subjects adapt their contribution levels closer to the others' average. Non-excluded subjects, who were already quite close to the group average, move closer or keep their levels constant.

previous phases increases the likelihood of being excluded again by about 11 percentage points in the B10 treatment. Having been excluded in the previous phase increases the likelihood of being excluded again by about 50 percent in the B8 treatment. This finding indicates that the adjustment of the excluded players is insufficient so that they face a higher risk of being excluded (again) than the non-excluded players.

Table B.6. Probability of being excluded when game B was played in previous and current phase

	(1) B10	(2) B10	(3) B8	(4) B8
Excluded in previous phase (d)	-0.0719		0.4973***	
	(0.0751)		(0.1649)	
Excluded before (d)		0.1125**		0.2599
		(0.0532)		(0.1786)
Voted for game B (d)	-0.0978	-0.0654	0.1122***	-0.0494
	(0.0855)	(0.0745)	(0.0347)	(0.0985)
Average group contribution in previous phase (%)	-1.0168***	-0.4134**	-0.3189*	-0.4890***
	(0.1562)	(0.1720)	(0.1656)	(0.1212)
Observations	200	215	55	75

Marginal or discrete effects from random effects probit estimation with standard errors in parentheses. Standard errors are clustered by group. Dependent variable is the probability of being excluded in the current phase when game B is played in both the current and the previous phase. *Average group contribution in previous phase* is defined as the contribution of the group averaged across all members and rounds of the previous phase in percent of endowment. Level of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimations include controls for phases. (d) indicates dummy variable.

Table B.7 shows whether there is a change in contribution behavior when subjects receive an exclusion vote but are not excluded in that round. For this analysis we focus only on the subjects who contributed less than the average contribution of their group members in the round of the vote. Subjects who receive a vote adapt their contribution closer to the average contribution level of the others in both treatments. This is also the case for low contributors who do not receive a vote for their exclusion—but their adjustment is much smaller.[11]

[11] The result also holds when we compare all subjects (irrespective of the contribution level) that receive a vote for their exclusion but are not excluded to those who do not receive a vote for their exclusion. The average increase in relative contribution is then smaller for subjects who receive a vote, because this sample also includes high contributors. Subjects who do not receive a vote decrease on average their contribution compared to the average contribution of their group members.

Table B.7 Individual contributions and gap between individual contribution and average contribution of other group members in percentage points for low contributors

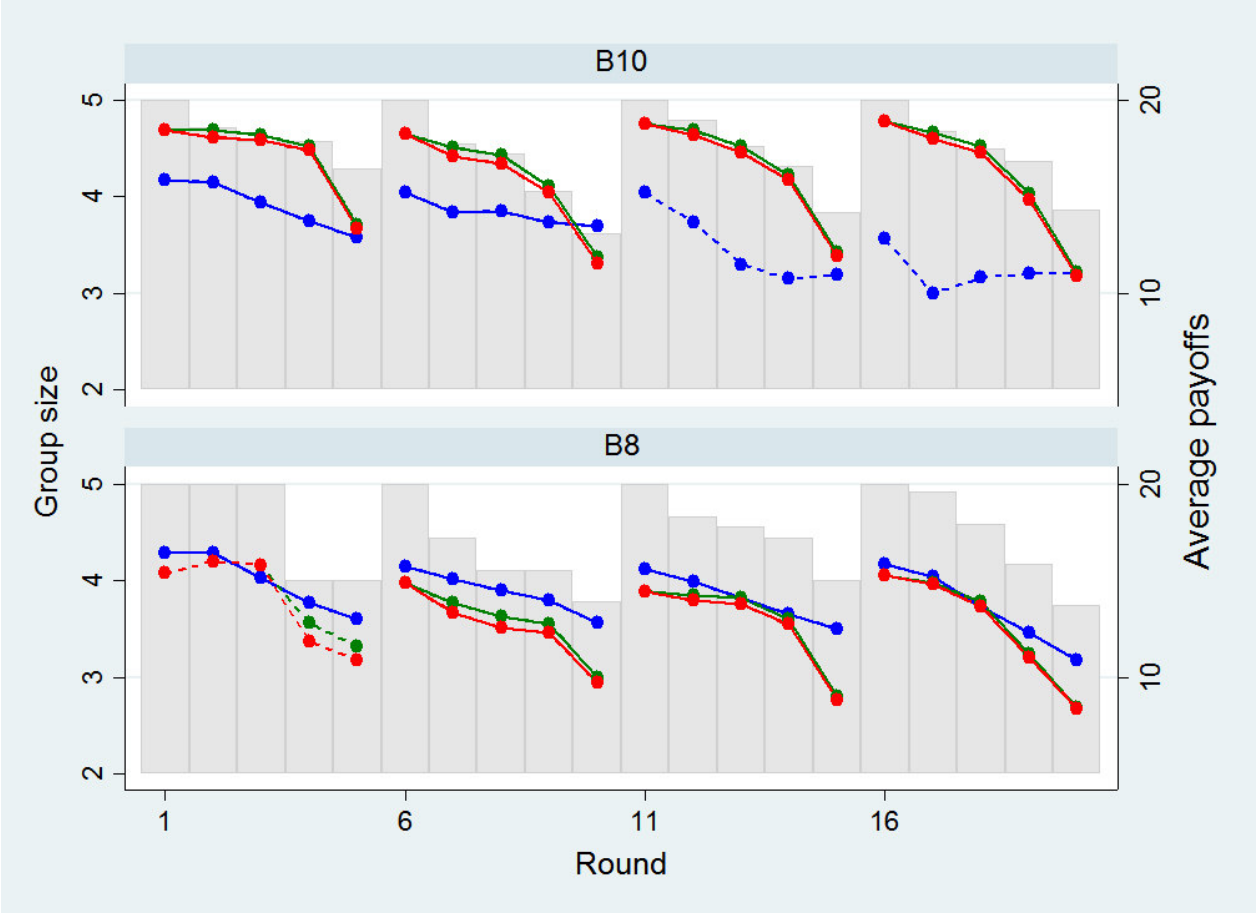
Treatment	Received vote in round t	Observations	Round	Contributions (%)		Gap in contributions (pp)	
				Mean	Std. Dev.	Mean	Std. Dev.
B10	Yes	80	t	56	33	-29	24
			t+1	57	43	-11	30
	No	85	t	59	29	-17	18
			t+1	62	38	-5	25
B8	Yes	33	t	39	34	-40	24
			t+1	50	43	0	25
	No	36	t	65	22	-12	14
			t+1	71	25	-3	17

The table shows individual contribution as percentage of endowment and the average gap between own contribution and average contribution of other group members in percentage points for subjects who did or did not receive a vote for their exclusion, but were not excluded, in the round of the vote and the following round. Round t does not include the last round of a phase and round t+1 does not include the first round of a phase. Only individuals who contributed less than the average contribution of their group members in round t are included in the analysis.

B.4 Payoffs

Figure B.2 shows average payoffs for each phase and distinguishes between non-excluded and excluded players in game B. By design, payoffs of the non-excluded players are higher on average than the payoffs of the whole group.

Figure B.2. Average payoffs by treatment



Average payoffs in game A (blue) and in game B (red). The green line shows average payoffs of non-excluded players in game B only. The dashed lines indicate that the data point is based on only few observations (N < 5). The bars show the average group size in the B-games.

Table B.8 shows average payoffs of the cooperators (the non-excluded players in game B and the four highest contributors in game A). The cooperators' payoff is always higher in game B than in game A when there is no institutional cost. In the B10 treatment, the difference is statistically significant for phase I and III (MWW test, $p < 0.1$ each). In the B8 treatment, the cooperators' payoffs are mostly lower in game B than in game A, but the differences are never statistically significant.

Table B.8 Average payoffs of cooperators

Phase	Game	B10	B8
I	A	13.8827 (2.9704, 16)	14.4336 (2.5028, 22)
	B	17.0138** (2.1729, 7)	14.1600† (-, 1)
II	A	13.5260 (4.2736, 5)	13.8900 (2.8367, 14)
	B	15.7183 (2.4899, 18)	12.6167 (1.9007, 9)
III	A	12.0775 (1.3471, 4)	13.5057 (12.7793, 14)
	B	16.3098† (2.5386, 19)	12.7793 (1.6130, 9)
IV	A	10.8400 (-, 1)	12.7285 (2.4225, 11)
	B	15.9182 (2.0875, 22)	12.4536 (1.6765, 12)
Total	A	13.4434 (2.9184, 16)	13.4913 (2.8182, 16)
	B	15.6749** (2.3097, 23)	12.3430 (2.4303, 23)

Single cells show mean of average payoffs of cooperators, conditional on treatment, phase, and game. Standard deviation and number of groups are given in parentheses. In game A the average excludes the lowest contributor(s) in the respective phase. In game B only non-excluded subjects are considered. Stars indicate significant difference between game A and game B in the same treatment and phase using the MWW test statistic. Level of significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. † indicates that the number of observations is too low to run a test ($N < 5$).

C. Experiment Instructions

These are the experimental instructions for the treatment *B8* (translated from German language). The instructions for the other treatments are very similar.

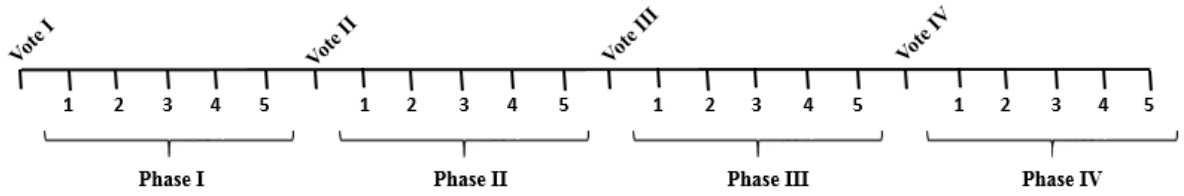
1. General information

In our experiment you can earn money. How much you earn will depend on the game play, or more precisely on the decisions you and your co-players make. For a successful run of this experiment, it is essential that you do not talk to other participants. Now, read the following rules of the game carefully. If you have any questions, raise your hand. We will come to you and answer them.

2. Game rules

There are five players in your group, meaning you and four other players. Each player is faced with the same decision problem. All decisions in the experiment are anonymous.

There are two games: game A and game B. At the beginning, every player in your group will vote for one of the two games. The game that receives the most votes (at least 3 out of 5) will be played by the group. The players will not be informed about the precise distribution of votes, but they will learn for which game most players have voted. The chosen game will be played five times (Phase I). After this, the group will vote a second time between game A and game B and play the chosen game another five times (Phase II). After this, the group will vote a third time and play the chosen game five times (Phase III). Finally, the group will vote a fourth time and play the chosen game five times (Phase IV). Hence, in total, your group will vote four times between game A and game B and play a total of twenty rounds, as indicated in the timeline below. You will play with the same group of players throughout all 20 rounds.



Game A works as follows: Each of the five players receives 10 tokens in the beginning of each round. The players decide if they keep their tokens or contribute them to a common project. The tokens that a player keeps benefit only that player. The tokens that a player contributes to the common project benefit all players. For each token that is contributed to the common project every player in the group will get 0.4 tokens. So, every player benefits from the tokens that have been contributed to the

common project regardless of how much they themselves have contributed. A player's profit is the sum of the tokens kept and the tokens that he or she receives from the common project.

Examples: If all five players keep their tokens and do not contribute any tokens to the common project, every player will get 10 tokens (= 10 + 0.4*0). If each of the five players contributes 10 tokens to the common project, each of them will get 20 tokens (= 0 + 0.4*50). If three players contribute 5 tokens each and two players contribute nothing, the former three players will get 11 tokens each (= 5 + 0.4*15) and the latter two players will get 16 tokens (= 10 + 0.4*15).

All five players decide simultaneously how much they contribute to the common project. Any integer amount between 0 and 10 tokens is possible. After all players have chosen their contributions to the common project, the contributions of all players will be shown on the screen. Here is an example for the presentation of the players' contributions:

Spielernummer	Beitrag in Taler	Gewinn in Taler
Spieler 1:	5	15.0
Spieler 2:	6	14.0
Spieler 3:	7	13.0
Spieler 4:	3	17.0
Spieler 5:	4	16.0
----------	----------	----------
Summe	25	75.0

Please note that the participant numbering is random and changes every round. Therefore, you and your co-players will not appear under the same number each round. Your own contribution will always be shown in red color. After the presentation of players' contributions, the round ends.

Game B works very similar, but there are two differences to game A. First, at the beginning of a round, every player receives 8 tokens (instead of 10 tokens). As in game A, the players decide simultaneously if they keep the tokens or contribute them to a common project. For each token that is contributed to the common project, every player in the group will get 0.4 tokens. A player's profit is the sum of the tokens kept and the tokens that he or she receives from the common project. As in game A, players' contributions will be displayed on the screen with randomized participant numbering in each round.

Examples: If all the players keep their tokens and do not contribute any tokens to the common project, every player will get 8 tokens (= 8 + 0.4*0). If every player contributes 8 tokens to the common project, each of them will get 16 tokens (= 0 + 0.4*40). If three players contribute 5 tokens each and

two players contribute nothing, the former three players will get 9 tokens each (= 3 + 0.4*15) and the latter two players will get 14 tokens (= 8 + 0.4*15).

The second difference is that there is an additional stage in game B. When the contributions to the common project are displayed, players can vote to exclude a member from the group. Every player can cast one vote to determine who should be excluded. Your vote can be cast by clicking on the corresponding box next to a participant's contribution. You cannot vote for yourself. It is possible not to cast a vote at all. To do so you can click the box next to "Nobody". Here you see an example:

Spielernummer	Beitrag in Taler	Gewinn in Taler	Mein Vorschlag
Spieler 1:	5	13.0	☐
Spieler 2:	6	12.0	☐
Spieler 3:	7	11.0	☐
Spieler 4:	3	15.0	☐
Spieler 5:	4	14.0	☐
Niemand			☐
----------	----------	----------	
Summe	25	65.0	

All players are informed about whether and how often they have been proposed for exclusion, but not by whom. A player who receives votes from more than half of his or her co-players will be excluded for the subsequent rounds of play. The exclusion, however, only prevails until the next choice of game A or game B. For example, if a player gets excluded by the co-players in round 3, he or she will be excluded from the group in rounds 4 and 5. After the fifth round, he or she will return to the group and the entire group will choose again between game A and game B. In round 5 it is therefore not possible to exclude any player.

By the exclusion of players it is possible that the group shrinks. If the group consists of five members, a player must receive at least 3 votes in order to be excluded. If the group consists of three or four members, a player must receive at least 2 votes in order to be excluded. If the group consists of only two members, exclusion is no longer possible. A summary of these exclusion rules is provided in the following table.

Current group size	Minimum number of votes that will lead to an exclusion of a player
5	3
4	2
3	2
2	No further exclusion possible

Excluded players receive precisely 8 tokens per excluded round. They cannot contribute to the common project and they do not receive any tokens from the common project. As long as they are excluded, they cannot cast a vote for another player to be excluded. They can, however, observe what happens in the game.

Your final payoff in the experiment is the sum of tokens you have earned across all 20 rounds. You will get 0.05 euros for each token. If, for example, you have earned 300 tokens across all rounds, you will get 15 euros at the end.