ORIGINAL PAPER

Voting on the threat of exclusion in a public goods experiment

Astrid Dannenberg[1] · Corina Haita-Falah[1] · Sonja Zitzelsberger[1]

Abstract

Ostracism is practiced by virtually all societies around the world as a means of enforcing cooperation. 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.

Data and program codes can be found here: https://doi.org/10.5281/zenodo.2594793.

✉ Astrid Dannenberg
dannenberg@uni-kassel.de

Corina Haita-Falah
corina.haita@gmail.com

Sonja Zitzelsberger
sonja.zitzelsberger@uni-kassel.de

1   Department of Economics, University of Kassel, 34117 Kassel, Germany

 Springer

1 Introduction

Cooperation among nonrelatives occurs frequently, for example among employees 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 government 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).

In this paper, we investigate a particular form of punishment, namely ostracism, in an experimental setting. Ostracism refers to the general process of excluding individuals from a group. We know from previous experiments that the option to exclude other players increases cooperation similar to other forms of punishment (see the next section). Our main interest in this paper is on whether people choose ostracism as a punishment institution when they have the choice and how this decision affects cooperation as compared to an exogenously imposed institution.

Ostracism 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 (Boehm 1986; Gruter 1986; Zippelius 1986; Gruter and Masters 1986). Imprisonment can be interpreted as modern version of ostracism executed by the government. Many groups that exist in modern democratic societies, like political parties, companies, or nonprofit associations, have rules that determine if and under what 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. 2002). 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).[1] 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,

---

[1] 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" (available at www.gruene.de/satzung, accessed January 2019). 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" (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).

banishment from the village (Ostrom 1990). Microfinance groups whose members borrow under joint liability often exclude individuals who fail to repay the loan from social activities (Baland et al. 2017; Putnam et al. 1994).

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 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). 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). 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.[2] Depending on their choice, groups either have the option to exclude members over the course of the game or not. Excluding a player necessarily means that the social optimum is no longer available as the group loses a potential contributor. We distinguish between a costless exclusion institution and a costly exclusion institution which, if chosen, reduces the endowment of all players. If there is an institutional cost, then implementing the institution forecloses achievement of the social optimum even if no group member is excluded. With this design, we want to test whether groups choose to implement the exclusion institution, how this choice affects cooperation, and how a fixed institutional cost affects the decisions and outcomes. We also compare an exclusion institution that is endogenously chosen by the groups to one that is 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 those with a long history, it is impossible to say when exactly members agreed to use ostracism as a way to punish defectors. In some cases, exclusion might have started as an ad-hoc reaction by a few members of the society and then developed

---

[2] 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.

into a social norm or tradition without ever being openly discussed and democratically chosen. Nevertheless, at any given point in time, the preservation and functioning of an institution depend on an internal agreement of at least some of the society's members. Curtailing the institution formation process into a limited number of decisions in a short period of time allows us 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, even when it comes at a cost. 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, but only when the institution is implemented. With respect to how the institution is implemented, we find that groups that choose the institution endogenously 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 which predicts that the threat of exclusion is not sufficient to support cooperation in a finitely repeated game. Given this inconsistency, 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 is social and the minority is selfish.

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 Sect. 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 of ESM) and Sect. 6 discusses the results and concludes.

2 Previous experimental literature

Numerous public goods experiments have shown that an option to reduce other players' payoffs increases contributions in public goods games (Fehr and Gächter 2000; for a review see Chaudhuri 2011). Despite higher contributions, however, overall payoffs do not necessarily increase due to the costs that occur on the side of the punishers and their targets. Clear payoff advantages are often realized only towards the end of the game or in games with 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).[3] 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. 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 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 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.

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 (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, 2018). But also weak institutions that do not change the nature of the game can have significant effects on cooperation and be quite popular (Feld and Tyran 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

---

[3] Kopányi-Peuker et al. (2018) study the effects of an exclusion institution in a weakest-link game. Croson et al. (2015) show that an automatic exclusion institution that always excludes the lowest contributor leads to very high cooperation levels.

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.

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 that change the equilibrium of the game, simply because strong institutions have a large effect on cooperation irrespective of how they are implemented. By contrast, the difference can be quite large for weak institutions that do not change the equilibrium (Dannenberg and Gallier 2019). We contribute to this literature by comparing the effects of an exclusion institution that is endogenously chosen and one that is exogenously imposed. This comparison has not been provided yet and it is interesting as the exclusion institution is weak from a theoretical perspective but it may be perceived as a relatively strong institution by the players.

3 Experimental design

3.1 The public goods 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).[4] 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 $i \in \{1, \ldots, n\}$ 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 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 the player's reputation formed in the course of the game, but only on his or her contribution in that round.

---

[4] 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) and Barrett and Dannenberg (2017).

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% 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. Thus, while blind revenge against the group is possible, targeted retaliation after exclusion is not. 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.[5] 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 interpreted that the community has the power to exclude individuals from the social benefits but not to take away their source of livelihood.

3.2 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.[6] The reduced endowment in B8 compared to game A can be interpreted as

---

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

[6] 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 C of ESM.
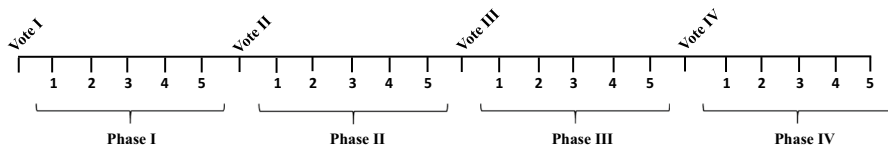
Fig. 1 Voting rounds and phases

a collective cost of the exclusion option. We set the fixed cost of the institution to 20% 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.

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 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).

3.3 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.[7] 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 the 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.

3.4 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*.[8] For the exogenous control treatments, we conducted ten sessions in September and November 2018 at the same computer lab and assigned them randomly to

---

[7] 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.

[8] 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).

B10-exo and *B8-exo*.[9] 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 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, subjects were paid their earnings privately in cash.

4 Theoretical background

In this section we derive the equilibria of games A and B, using standard preferences and two models of social preferences. For each model, we provide predictions for players' choice between the games, assuming common knowledge of preferences throughout. For brevity we present only a summary of the main results, while the complete analysis with the proofs can be found in Appendix A of ESM.

4.1 Standard preferences model

In the standard preferences model, zero contribution by all players is the unique Nash equilibrium (NE) of the stage game. 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. By backward induction it obtains that the unique subgame perfect Nash equilibrium (SPNE) of the repeated game is zero contribution by all players in each round, regardless of the game played. Thus, players are indifferent between game A and game B10, but prefer A to B8 as the former gives a higher endowment and so a higher payoff. Hence, the standard preferences model predicts that game B8 is never played when the choice is between A and B8. When the choice is between A and B10, each game will be played half the time. If B10 is chosen, then any configuration of votes and group sizes can be part of an equilibrium

---

[9] 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.

because exclusion in our setting is costless and players are thus indifferent between excluding and not excluding a group member (see Appendix A.1 of ESM).

4.2 Inequality aversion model

In the inequality aversion model by Fehr and Schmidt (1999), players derive utility from the material earnings resulting from the public good, and they 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 other group members (disadvantageous inequality aversion). Specifically, the inequality averse utility function 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 good, α_i measures the aversion to disadvantageous inequality and $0 \leq \beta_i < 1$ captures the aversion to advantageous inequality. Moreover, $\alpha_i \geq \beta_i$ such that players are more averse to disadvantageous than to advantageous inequality.

With these 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. $\beta_i \geq 1 - a = 0.6$. We call these players conditional cooperators, following the original paper. This equilibrium exists in both games and it makes no use of the exclusion option in game B. However, it requires coordination on a certain contribution level for which full contributions seems to be a natural focal point as it is Pareto dominant. By backward induction it obtains that full contributions in each round is a SPNE, 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 groups of inequality averse players are indifferent between playing A and playing B10, but they strictly prefer A to B8 (see Appendix A.2 of ESM).

With one selfish player in the group, with $\beta_i < 1 - a = 0.6$, the unique equilibrium of the stage game is zero contribution by all players, since this strategy is dominant for the selfish player.[10] Given this, it is also the best response of the remaining conditionally cooperative players, i.e. those players for which $\beta_j \geq 0.6$.[11] However, in game B, the conditional cooperators can use the exclusion institution against 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. Because exclusion is not possible in game A, the only SPNE of game A is zero

---

[10] The situation with two selfish players is qualitatively similar (see Appendix A.2 of ESM).

[11] Our parameter values satisfy the condition from Proposition 4, Part b in Fehr and Schmidt (1999).

contribution by all players in every round. Given these equilibrium outcomes, the selfish player either strictly prefers game A over both B8 and B10 (if $\alpha_i > 0$) or she prefers A over B8 and is indifferent between A and B10 (if $\alpha_i = 0$).[12] 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. With our experimental parameters, these players should contribute more than 5 tokens for game B8 to be preferred (see Appendix A.2 of ESM).[13]

4.3 Reciprocity model

The reciprocity model of Rabin (1993) assumes that people derive utility from reciprocation of kindness and unkindness, in addition to the monetary gains. 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. $\beta_i = \beta > 2E_p(1-a)$, $\forall i = 1, \ldots, n$, then the stage game has two pure-strategy NE, one in which all players contribute zero and one in which all players contribute their full endowment. For the SPNE, in which either of the stage-game equilibria is repeated every round, the symmetry of the equilibrium leaves the exclusion institution in game B unused (or ineffective if used in the zero-contribution equilibrium). Hence, groups of highly reciprocal players are indifferent between A and B10, but prefer A to B8 due to the higher endowment (see Appendix A.3 of ESM). If players are not sufficiently reciprocal, i.e. $0 < \beta \leq 2E_p(1-a)$, then zero contribution by all players is the only equilibrium.

---

[12] 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 the second round 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 of ESM.

[13] 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.

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 NE. The non-reciprocal player contributes zero, regardless of what the reciprocal players do. Apart from the equilibrium in which all players contribute zero, there 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 repetition of either of the two pure-strategy NE is a SPNE of game A. However, in game B, the SPNE that involves the full-contribution equilibrium by the highly reciprocal players in all rounds 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.

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 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 of ESM).

4.4 Differences between standard and social preferences

In summary, 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. When social preferences are assumed, 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 are crucial for the game choice. Groups consisting solely of social individuals can sustain cooperation in both games. If at all, they choose the exclusion institution only if it is costless and they do 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 implement the exclusion institution, exclude the selfish player from the group, and cooperate thereafter. With some restrictions, this is also true when the exclusion institution is costly. While it is not our intention to test the two theories of social preferences with this experiment, as has been done for example by Blanco et al. (2011), we use them to provide possible explanations for why players may vote for and use the exclusion institution.
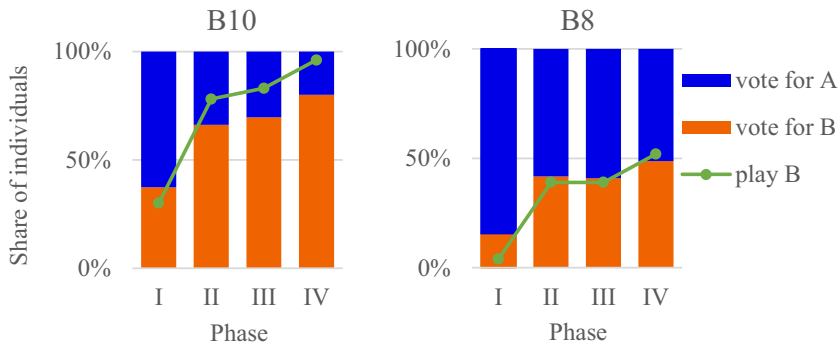
Fig. 2 Voting behavior and game choice by treatment

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 B of ESM.

5.1 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 votes 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% in the first phase to 80% in the last phase and the share of groups that play B rises from 30% to 96%. In *B8*, the share of B-voters increases from 16% to 49% and the share of groups that play B increases from 4% to 52%. 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).[14]

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% of the individuals who start the first phase by

---

[14] 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.

Table 1 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

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% 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% consistently votes for B without switching to A. Likewise, at the group level, 75% of groups in B10 and 68% 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 1 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.

5.2 Contributions

Table 2 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% in game A and 76% in game B. In *B8,* the average contribution is 41% in game A and 74% 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.[15] 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 of ESM). Thus, the exclusion of low contributors alone cannot explain the higher average contributions in game B.

5.3 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% in the first round of playing game B while B-voters contribute 83% in

---

[15] 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% of groups, the lowest contributor is a single player. In 2% 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.

Table 2 Average contributions and payoffs by treatment, phase, and game

		B10					B8				
Phase	Game	Number of groups in each game	Average group size	Average contribution (in %)	Average payoff (insider)	Average payoff (all)	Number of groups in each game	Average 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	93.00ᵃ	14.5ᵃ	14.0ᵃ
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	76.19ᵃ	16.9ᵃ	16.4ᵃ	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	72.42ᵃ	16.5ᵃ	16.0ᵃ	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: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

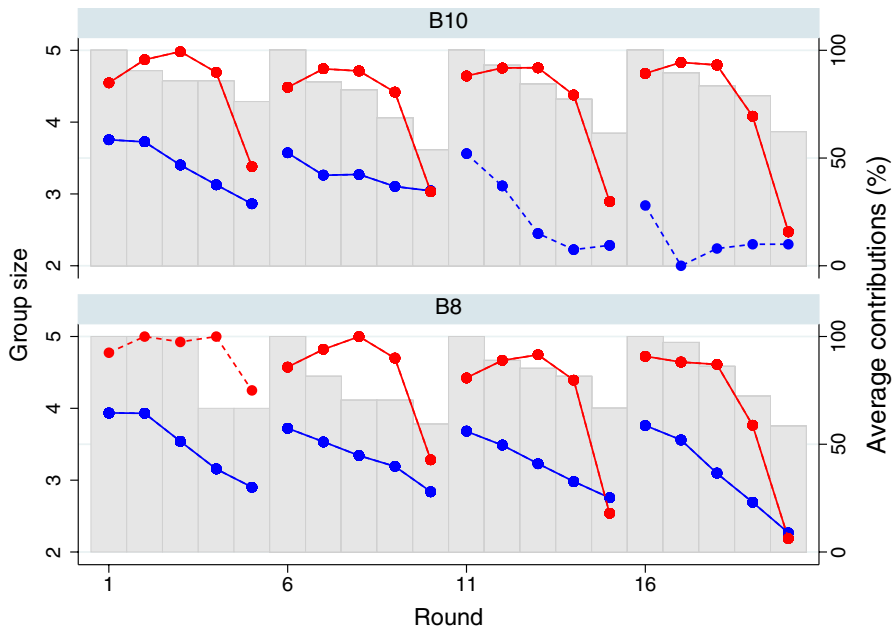ᵃIndicates that the number of observations is too low to run a test (N < 5)

Fig. 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

the first round. In *B8*, A-voters contribute on average 63% in the first round while B-voters contribute 87%. Table 3 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 of ESM (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.

Columns (5)–(8) in Table 3 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% in the first round while B-voters contribute 61% in the first round. In *B8*, A-voters contribute 66% in the first round and B-voters contribute 64%. The corresponding regression

Table 3 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 (%) in previous phase		0.3874***		0.5416***		0.1320*		0.2285
		(0.1118)		(0.0832)		(0.0702)		(0.1468)
Game B in previous phase (d)						0.0495		−0.1431
						(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

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 of ESM (Tables B.3 and B.4). The regressions shown in Table 3 and in the Appendix of ESM 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.

5.4 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% of groups keep a group size of five throughout the phase, 47% reach a group size of four, 17% a group size of three, and 12% a group size of two. The respective numbers for the *B8* treatment are 26, 42, 23, and 10%.

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% in *B10* and 76% in *B8* receive at least one vote for their exclusion. Forty-seven percent in *B10* and 37% 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 of ESM). 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

⚛ Springer

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 of ESM).

5.5 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 2 (and Figure B.2 in the Appendix of ESM) show that, in *B10*, average payoffs in all phases are higher in game B than in game A and the difference 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 of ESM 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.

5.6 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 contribution 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 $p = 0.1814$). 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,
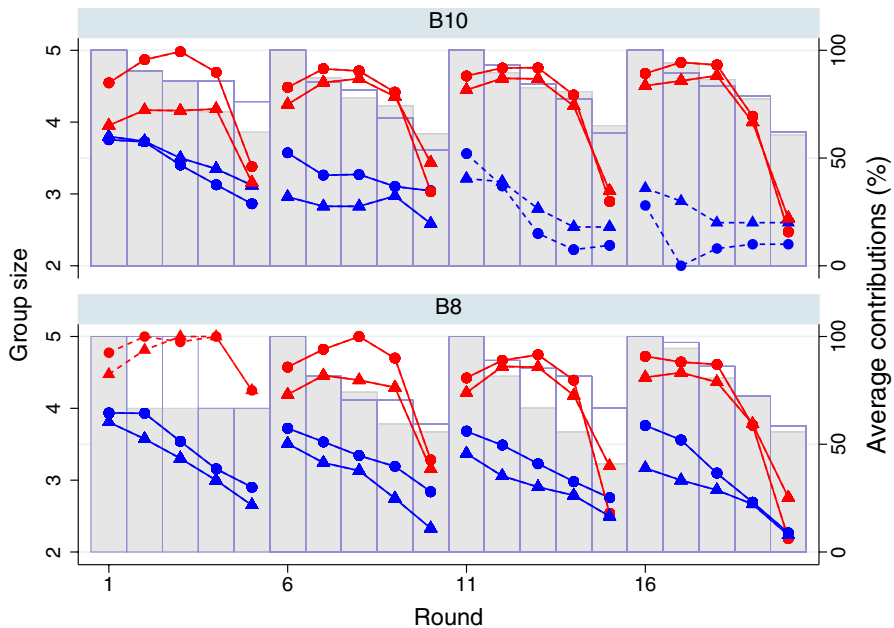
Fig. 4 Comparison of contribution rates between endogenous and exogenous treatments. The figure shows average contributions over time, measured in percent of endowment, in game A (blue) and in game B (red) by treatment. The dashed lines indicate that data points are based on only few observations ($N < 5$). Observations in the exogenous treatments are marked by triangles. Observations in the endogenous treatments 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)

$p < 0.1$ each). In *B8-exo*, average payoffs are higher in the A game than the B game and the difference is 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 Fig. 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 in *B8-exo* and the difference is weakly significant in phase III ($p = 0.0965$). Regarding exclusions of individuals and exclusion votes, we find no significant differences between the endogenous and the corresponding exogenous games ($p > 0.1$ each).[16]

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

---

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

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% in *B10-exo* and 71% in *B8-exo*) is slightly lower than the 80% found by Maier-Rigaud et al. (2010) and the 90% 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.

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 can also test the differences 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 have been excluded or received a vote for exclusion adjust 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 social preferences models. 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. 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.

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 the 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.

Springer

References

Akpalu, W., & Martinsson, P. (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, J.-M., Gangadharan, L., Maitra, P., & Somanathan, R. (2017). Repayment and exclusion in a microfinance experiment. *Journal of Economic Behavior & Organization, 137,* 176–190.

Barrett, S., & Dannenberg, A. (2017). Tipping versus cooperating to supply a public good. *Journal of the European Economic Association, 15*(4), 910–941.

Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior, 72*(2), 321–338.

Boehm, C. (1986). Capital punishment in tribal montenegro: Implications for law, biology, and theory of social control. *Ethology and Sociobiology, 7*(3–4), 305–320.

Bolleyer, N. & Gauja, A. (2015). Legal conceptions of organizational membership: Implications for intra-party dynamics and democracy. Paper Prepared for the Political Studies Association (UK) Annual Conference 2015.

Brekke, K. A., Hauge, K. E., Lind, J. T., & Nyborg, K. (2011). Playing with the good guys. A public good game with endogenous group formation. *Journal of Public Economics, 95,* 1111–1118.

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

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

Cinyabuguma, M., Page, T., & Putterman, L. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics, 89*(8), 1421–1435.

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

Dal Bó, P. (2014). Experimental evidence on the workings of democratic institutions. In S. Galiani & I. Sened (Eds.), *Institutions, property rights, and economics growth: The legacy of Douglass North* (pp. 266–288). Cambridge: Cambridge University Press.

Dal Bó, E., Dal Bó, P., & Eyster, E. (2018). The demand for bad policy when voters underappreciate equilibrium effects. *Review of Economic Studies, 85*(2), 964–998.

Dal Bó, P., Foster, A., & Putterman, L. (2010). Institutions and behavior: Experimental evidence on the effects of democracy. *American Economic Review, 100*(5), 2205–2229.

Dannenberg, A., & Gallier, C. (2019). The choice of institutions to solve cooperation problems: A survey of experimental research, MAGKS Discussion Paper Series No. 11-2019.

Davis, B. J., & Johnson, D. B. (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. CIRANO Scientific Series No. 99s-24. Available at https://depot.erudit.org/bitstream/000355dd/1/99s-24.pdf.

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

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994.

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

Fehr, E., & Williams, T. (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, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological Science, 25*(3), 656–664.

Feld, L. P., & Tyran, J.-R. (2002). Tax evasion and voting: An experimental analysis. *KYKLOS*, *55*, 197–222.

Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 271–279.

Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science, 322,* 1510–1511.

Gruter, M. (1986). Ostracism on trial: The limits of individual rights. *Ethology and Sociobiology, 7*(3–4), 271–279.

Gruter, M., & Masters, R. D. (1986). Ostracism as a social and biological phenomenon: An introduction. *Ethology and Sociobiology, 7*(3–4), 149–158.

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science, 312,* 108–111.

Kamei, K., Putterman, L., & Tyran, J.-R. (2015). State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods. *Experimental Economics*, *18*(1), 38–65.

Kopányi-Peuker, A., Offerman, T., & Sloof, R. (2018). Team production benefits from a permanent fear of exclusion. *European Economic Review, 103,* 125–149.

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

Markussen, T., Putterman, L., & Tyran, J.-R. (2014). Self-organization for collective action: An experimental study of voting on sanction regimes. *Review of Economic Studies, 81*(1), 301–324.

Masclet, D. (2003). Ostracism in work teams: A public good experiment. *International Journal of Manpower, 24*(7), 867–887.

Nyborg, K. (2017). Reciprocal climate negotiators. *Journal of Environmental Economics and Management, 92,* 707–725.

Ostrom, E. (1990). *Governing the commons. The evolution of institutions for collective action*. Cambridge: Cambridge University Press.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review, 86*(2), 404–417.

Page, T., Putterman, L., & Unel, B. (2005). Voluntary association in public goods experiments: Reciprocity, mimicry, and efficiency. *Economic Journal, 115,* 1032–1053.

Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1994). *Making democracy work: Civic traditions in modern Italy*. Princeton: Princeton University Press.

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

Solda, A., & Villeval, M. C. (2018). Exclusion and reintegration in a social dilemma. *Economic Inquiry*. https://doi.org/10.1111/ecin.12720.

Sutter, M., Haigner, S., & Kocher, M. G. (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. (2002). *National labour relations in internationalized markets. A comparative study of institutions, change, and performance*. Oxford: Oxford Univ. Press.

Tyran, J.-R., & Feld, L. P. (2006). Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics, 108*(1), 135–156.

Zippelius, R. (1986). Exclusion and shunning as legal and social sanctions. *Ethology and Sociobiology, 7*(3–4), 159–166.

Springer