# 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 analyses** – 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:

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

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 $\pi_i$ 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: $\alpha_i$, aversion to disadvantageous inequality and $\beta_i$, aversion to advantageous inequality.

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

<u>*Game A*</u>

<u>Case 1:</u> 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:

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

<u>Case 2:</u> 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,

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

for any $\beta_k \ge 0.6$ 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.

<u>*Game B*</u>

<u>Case 1:</u> For simplicity, let us first assume that there is only one selfish player with $\beta_i < 0.6$ and $n-1$ conditional cooperators with $\beta_j \ge 0.6, j \ne i$.[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:

$$U_j(\text{NOT exclude before the 5th round}) =$$

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

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

$$\begin{aligned} U_j(\text{exclude before round 5}) = \underbrace{4\times\left(E_B + (na-1)g\right)}_{\text{5 players in rounds 1 to 4}} + \underbrace{E_B + ((n-1)a-1)g - \frac{1}{n-1}\beta_j((n-1)a-1)g}_{\text{player } i \text{ excluded in the 5th round}} = \\ 5E_B + 4(na-1)g + g((n-1)a-1)\left(1 - \frac{1}{n-1}\beta_j\right) \end{aligned} \tag{A.2.8}$$

[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.

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

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

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

(A.2.9)

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

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

(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 $0 \leq g_i < g$. 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) =

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

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) =

$$3\times\left(E_B+(na-1)g\right)+\underbrace{E_B+((n-1)a-1)g-\frac{1}{n-1}\alpha_j g}_{\text{round4}}+\underbrace{E_B+((n-1)a-1)g-\frac{1}{n-1}\beta_j((n-1)a-1)g}_{\text{round5}}$$

(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 $\beta_j \leq \alpha_j$.

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:

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

<u>Case 2:</u> 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.

---

[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.

**Table A.2.1 Equilibrium utilities**

| | Game A | Game B |
|---|---|---|
| Case 1: $\exists \beta_i < 0.6$ | $U_i^A = 5E_A, \forall i$ | $U_i^B = 5E_B - 4((n-1)a-1)\alpha_i g$<br>$U_j^B = 5E_B + 4((n-1)a-1)\left(1-\frac{\beta_j}{n-1}\right)g$ |
| 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 $\beta_j$ 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 $\pi_i$ is the material payoff given by (A.1). Parameter $\beta_i$ 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 $G_{-i} = \sum_{l=1, l \neq i}^{n} g_l$. 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*:

[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.

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

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 $j=1,\ldots,i-1,i+1,\ldots,n.$ 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:

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

For $\frac{1}{E_p}\frac{G_{-i}}{n-1}=\frac{1}{2}+\frac{E_p(1-a)}{\beta}$ 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 $\beta \le 2E_p(1-a)$, 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

---

[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 $\beta$ of the cooperating players is low enough. However, this is not consisted with the value of $\beta$ 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

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.

*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 $\pi_{ij}^{max}$ and $\pi_{ij}^{min}$ are given by equations (A.3.1) and (A.3.2), respectively and $\pi_{ij}^{e}$ 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):

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

---

excluded per round. We find that this complication is not worth pursuing in order to get the gist of the game incorporating reciprocal preferences.

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:

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

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:

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

$(ii)$ $g_k = 0$ and $g_i = E_p,\ \forall i \neq k,\ i = 1,\dots,n$ 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 $j \in \{1,\dots,i-1,i+1,\dots,n-1\},\ j \neq k$. Hence, the reciprocal utility of player $i \neq k$ is:

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

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 $\beta_{excl.} \leq \beta \leq \beta_{not\ excl.}$, then the full contribution equilibrium exists only if the non-reciprocal player is excluded. Finally, if $\beta < \beta_{excl.}$, 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.

<u>*Game A*</u>

[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).

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:

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

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

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

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

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. $\beta > \dfrac{5a(n-1)^2(E_A - E_{B8})}{6}$. 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, Ernst and Klaus M. Schmidt (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* 92, 707-725.

Rabin, Matthew (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.

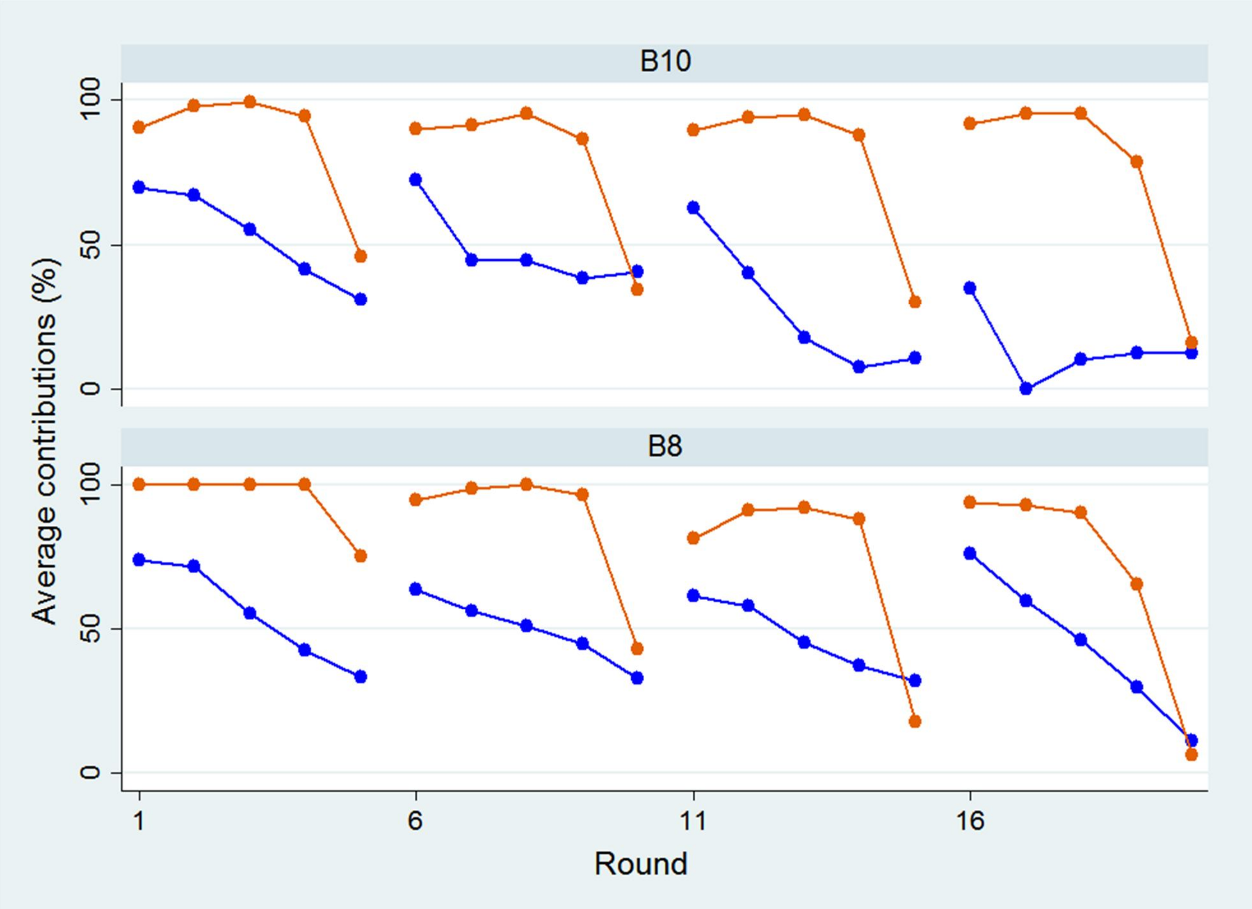
**Table B.1 Average contributions of cooperators**

| Phase | Game | B10 | B8 |
|---|---|---|---|
| I | A | 0.5282<br>(0.2804, 16) | 0.5523<br>(0.2403, 22) |
| | B | 0.8575**<br>(0.0429, 7) | 0.9500†<br>(-, 1) |
| II | A | 0.4810<br>(0.3601, 5) | 0.4950<br>(0.2829, 14) |
| | B | 0.7948*<br>(0.1395, 18) | 0.8640***<br>(0.0704, 9) |
| III | A | 0.2763<br>(0.1465, 4) | 0.4671<br>(0.2622, 14) |
| | B | 0.7930†<br>(0.1190, 19) | 0.7404***<br>(0.1205, 9) |
| IV | A | 0.14<br>(-, 1) | 0.4442<br>(0.2265, 11) |
| | B | 0.7536†<br>(0.1094, 22) | 0.6966***<br>(0.1245, 12) |
| Total | A | 0.4777<br>(0.2754, 16) | 0.4638<br>(0.2144, 22) |
| | B | 0.7725***<br>(0.1134, 23) | 0.7537***<br>(0.1184, 15) |

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) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | B10 | B10 | B8 | B8 | B10 | B10 | B8 | 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) | (2) | (3) | (4) | (5) | (6) |
| | B10 | B8 | B10 | B10 | B8 | 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 (%) Mean | Contributions (%) Std. Dev. | Gap in contributions (pp) Mean | Gap in contributions (pp) 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)<br>B10 | (2)<br>B10 | (3)<br>B8 | (4)<br>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**

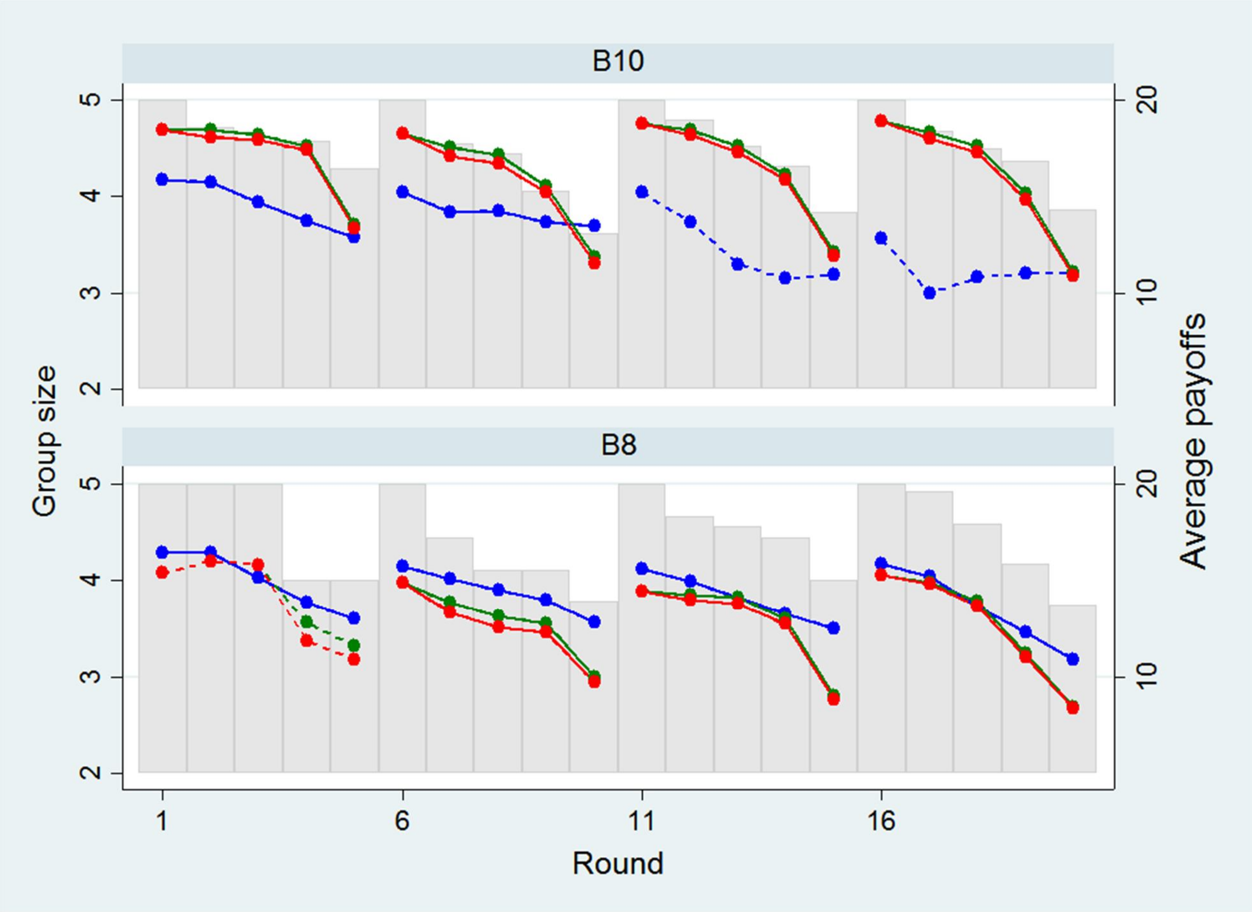
| Treatment | Received vote in round t | Observations | Round | Contributions (%) | | Gap in contributions (pp) | |
|---|---|---|---|---|---|---|---|
| | | | | Mean | Std. Dev. | Mean | Std. Dev. |
| **B10** | Yes | 80 | t | 56 | 33 | -29 | 24 |
| | | | t+1 | 57 | 43 | -11 | 30 |
| | No | 85 | t | 59 | 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 only marginally statistically significant when all phases are considered (MWW test, $p<0.1$).

## Table B.8 Average payoffs of cooperators

| Phase | Game | B10 | B8 |
|---|---|---|---|
| I | A | 13.8827<br>(2.9704, 16) | 14.4594<br>(2.4908, 22) |
| | B | 17.0138**<br>(2.1729, 7) | 14.1600†<br>(-, 1) |
| II | A | 13.5260<br>(4.2736, 5) | 13.8900<br>(2.8367, 14) |
| | B | 15.7183<br>(2.4899, 18) | 12.6167<br>(1.9007, 9) |
| III | A | 12.0775<br>(1.3471, 4) | 13.5057<br>(12.7793, 14) |
| | B | 16.3098†<br>(2.5386, 19) | 12.7793<br>(1.6130, 9) |
| IV | A | 10.8400<br>(-, 1) | 12.7285<br>(2.4225, 11) |
| | B | 15.9182<br>(2.0875, 22) | 12.4536<br>(1.6765, 12) |
| Total | A | 13.4434<br>(2.9184, 16) | 13.5042*<br>(2.1293, 22) |
| | B | 15.6749**<br>(2.3097, 23) | 12.3430<br>(1.6339, 15) |

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.

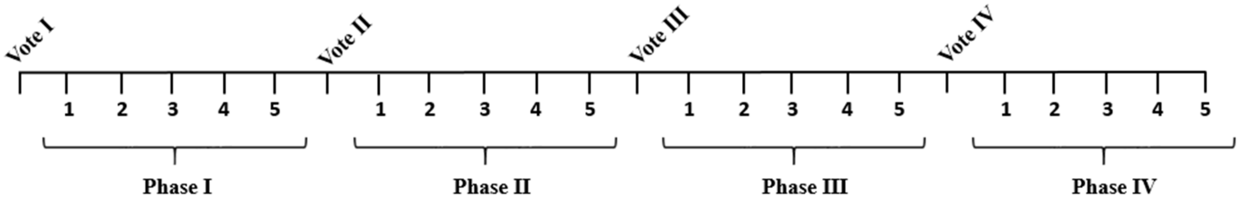
### *1. General information*

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.