Julian Detemple | Michael Kosfeld

Fairness and Inequality in Institution Formation

SAFE Working Paper No. 402 | September 2023

Leibniz Institute for Financial Research SAFE
Sustainable Architecture for Finance in Europe
info@safe-frankfurt.de | www.safe-frankfurt.de

Fairness and Inequality in Institution Formation

Julian Detemple and Michael Kosfeld*

August 24, 2023

Abstract

A key solution for public good provision is the voluntary formation of institutions that commit players to cooperate. Such institutions generate inequality if some players decide not to participate but cannot be excluded from cooperation benefits. Prior research with small groups emphasizes the role of fairness concerns with positive effects on cooperation. We show that effects do not generalize to larger groups: if group size increases, groups are less willing to form institutions generating inequality. In contrast to smaller groups, however, this does not increase the number of participating players, thereby limiting the positive impact of institution formation on cooperation.

Keywords: Institution formation, group size, social dilemma, social preferences
JEL Classification: C92, D02, D63, H41

*Faculty of Economics and Business, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany. E-mail: detemple@econ.uni-frankfurt.de, kosfeld@econ.uni-frankfurt.de. We thank seminar participants at Goethe University for helpful comments. We gratefully acknowledge support from the Experiment Center at the Leibniz Institute for Financial Research SAFE. Declarations of interest: none.

> Each society learns to live with a certain amount of such dysfunctional or mis-behavior; but lest the misbehavior feed on itself and lead to general decay, society must be able to marshal from within itself forces which will make as many of the faltering actors as possible revert to the behavior required for its proper functioning.
>
> Hirschman (1970, p.1)

A key solution to manage collective action problems in public good provision is the formation of institutions that commit players to cooperate (Baland and Platteau, 1996; Ostrom, 1999). Examples range from international agreements to halter climate change (e.g., Kyoto Protocol, Paris Climate Agreement) to the governance of common pool resources (e.g., Tang, 1992; Hviding and Baines, 1994; Agrawal and Yadama, 1997). Yet, in case only a subset of players agree to commit, such institutions create inequality as they give players who have not agreed to commit the possibility to free ride.

Previous lab experimental research has shown that such inequality affects the institution formation process, because players who care about inequality, or reciprocity, are unwilling to agree on commitments in the presence of free riders (Kosfeld, Okada and Riedl, 2009; see Dannenberg and Gallier, 2020 for a review). To provide a real-world example, debates along these lines emerged around the Kyoto Protocol and the Paris Climate Agreement once the U.S. had decided to quit the agreement. Experimental results suggest that as a consequence of these fairness concerns only *grand institutions*, i.e., agreements where *all* players are committed, are formed. Fairness concerns – or more generally social preferences – thus boost the positive effects of institutions on public good provision. The results are based, however, on relatively small groups of typically around four players, which raises the important question to what extent the results also hold in groups that are larger. Obviously, in most applications (cf. examples above) significantly more players negotiate an agreement. In this paper, we study the effects of a moderate increase in group size by doubling and tripling the number of players from four to eight and twelve and find that social preferences, instead of boosting efficiency, impede the positive effects of institutions on public good provision.

There exist at least three arguments why an increase in group size may influence the institution formation process in the presence of social preferences. Firstly, results from, for example, Schumacher et al. (2017) and Alós-Ferrer, García-Segarra and Ritschel (2022), albeit in different empirical settings, suggest that the role of social preferences becomes weaker when groups become larger. This implies that institutions smaller than the grand institution are more likely to be formed despite the inequality they entail. Secondly and in concordance with this, our theoretical results in this paper show that individual concerns from social preferences need to be increasingly strong for non-grand institutions to be rejected if groups become larger – even if social preferences themselves are not affected by group size. Again, the implication is that the likelihood for non-grand institutions to be formed should increase. Thirdly and arguing in the other direction, for any institution generating inequality to be blocked, there need to be players who are willing to veto. If social preferences continue to play a role for at least some players, the probability for this to happen mechanically increases with group size, thereby lowering chances for non-grand institutions to be implemented. Existing research is silent on which of the three effects will dominate when group size increases.

We analyze these questions in a stylized laboratory setting where, to the best of our knowledge, we are the first to exogenously vary group size in the context of institution formation. The controlled lab environment allows us to isolate important details of the institution formation process from other channels that are likely to be correlated with institutional outcomes in the field. For example, groups with more cooperatively minded members might be more successful in making use of institutions (Rustagi, 2022) and therefore more likely to endure, leading to a higher likelihood to appear in a sample. Theoretically, we build on the institution formation game as proposed and analyzed in Okada (1993) and Kosfeld, Okada and Riedl (2009). In this game, players can form and join institutions in which they are committed to contributing all their endowment to a linear public good. The formation of an institution is decided unanimously by all players who are willing to commit themselves.[1] Players who do not join are not affected by any commitments and can freely decide on their contribution. Still, they benefit from the other players' contributions. We modify the game such that the profitability of and the payoff inequality in an institution of the same *relative* size are independent of group size. In our theoretical analysis (cf. Appendix E), we show that as a result of this, the set of equilibria as well as the relative size of the unique strict institutional equilibrium are independent of group size if all players have money-maximizing preferences. If (some) players have social preferences, which we

[1]Both practical and theoretical reasons favor the unanimity rule over the majority rule. Practically, the lack of (existing) institutions to force people to cooperate lies at the heart of many collective action problems. Theoretically, the unanimity rule ensures the existence of a unique equilibrium in terms of institution size (cf. Appendix E).

model using the framework of Fehr and Schmidt (1999), we establish conditions such that the grand institution is a strict institutional equilibrium or even the unique institutional equilibrium of the game. Importantly, we show that these conditions on social preferences become more demanding (in terms of inequality to be disliked) if group size increases. Thus, social preferences, while continuing to favor the grand institution, need to be increasingly strong for players to reject non-grand institutions in larger groups. But having only one player with sufficiently strong social preferences in a group is sufficient to veto an institution. The likelihood of this increases with group size but decreases when social preferences play a weaker role. In sum, the analysis suggests that (i) if players don't have social preferences, increasing group size has no effects; (ii) if social preferences do play a role, the overall effect on the likelihood for non-grand institutions to be formed depends on how exactly group size affects the role of social preferences. Our design allows us to analyze this role precisely and compare its effects across group sizes. We compare three treatments in a between-subject design: small groups with four players, medium-sized groups with eight players, and large groups with twelve groups.

We report three main results from the experiment. First, increasing group size drastically *lowers* the probability that institutions smaller than the grand institution are implemented. For example, if 11 out of 12, or 7 out of 8, players are willing to form an institution, these institutions are significantly less likely to be implemented in the aggregate than institutions with 3 out of 4 players. Larger groups therefore favor the grand institution even more than smaller groups of four players. Second, this effect is driven by a higher likelihood that at least one player in the group has sufficiently strong social preferences to block any institution involving inequality. In fact, social preferences play a role as important in groups of four as in group of twelve players. In other words, group size does not affect social preferences themselves. For this, we show that individual voting decisions about institutions involving inequality are not affected by group size, neither when we compare the largest non-grand institution across group sizes nor when we consider the same relative group size. Instead we find that the (mechanical) increase in the likelihood of having one player who always votes against institutions involving inequality in a group explains the variation in implementation rates. Third, we show that the stronger emphasis on the grand institution does not translate into more successful formation of such institutions: when group size increases beyond four, average implementation rates fall below 15% and average efficiency decreases by more than 47%. This is caused by a decrease in the (relative) number of players willing to form an institution. Importantly, all results hold if we account for learning.

These findings have important implications. Group size severely affects the dynamics which are crucial for the success of institution formation in the presence of inequality, leading to lower efficiency in larger groups. Prior results based on small groups obviously cannot be generalized to scenarios involving as few as eight players or more. Moreover, our results suggest that social preferences still play an important role in larger groups: groups with as many as twelve players do not reveal more money-maximizing behavior than groups of four. However, without additional instruments to harness the positive effects of social preferences in this context, social preferences are a double-edged sword and hinder rather than support the realization of efficient outcomes by means of institution formation in larger groups. Additional instruments to both mobilize as many players as possible to join and reach a consensus among players who are willing to join on how much inequality to tolerate seem to be necessary.

Our paper relates to several strands in the literature. First, our findings contribute to the analysis of heterogeneity between players in the provision of public goods (e.g., Cherry, Kroll and Shogren, 2005; Fischbacher and Gächter, 2010; de Oliveira, Croson and Eckel, 2015) and institutions aimed at promoting cooperation in collective action problems (e.g., Poteete and Ostrom, 2004; Dannenberg, Lange and Sturm, 2014; Gangadharan, Nikiforakis and Villeval, 2017; Schmidt and Ockenfels, 2021). Closest to our study is work by Schmidt and Ockenfels (2021), who show that in groups of four players commitment institutions are superior to individual commitments in the presence of heterogeneity in endowment and how much players benefit from the public good. We contribute to this literature by highlighting that even when players are identical on observables (e.g., endowment), heterogeneity in (unobservable) preferences can still impact the ability of institutions to solve collective action problems in larger groups. While we consider heterogeneity in social preferences, our results are also informative about other factors which could give rise to heterogeneity in commitment preferences.

Next, our results contribute to the literature on institution formation in lab experiments (e.g., Page, Putterman and Unel, 2005; Ertan, Page and Putterman, 2009; Sutter, Haigner and Kocher, 2010; Markussen, Putterman and Tyran, 2014; Kamei, Putterman and Tyran, 2015; Dickinson, Masclet and Villeval, 2015; Kamei, Putterman and Tyran, 2019; Schmidt and Ockenfels, 2021). While groups with more than four players have been studied (e.g., Gürerk, Irlenbusch and Rockenbach, 2006, 2014; Dannenberg, Lange and Sturm, 2014; McEvoy et al., 2011; Kamei, Putterman and Tyran, 2019), we are – to the best of our knowledge – the first to analyze group-size effects by exogenously varying the size of the group in the context of institution formation in social dilemma situations. Moreover, we do so in an environment in which the material incentives to form and the inequality in an institution of the same relative size are *orthogonal* to group size. As spelt out above, this allows us to trace the overall negative effect of group-size on efficiency to how group size affects the very drivers of institution formation which this literature has highlighted (Kosfeld, Okada and Riedl, 2009; Dannenberg and Gallier, 2020).

We also contribute to the literature on punishment in social dilemma situations. This literature has stressed the importance of decentralized peer punishment in achieving co-operative outcomes (Fehr and Gächter, 2000, 2002; Carpenter, 2007; Masclet and Villeval, 2008; Ambrus and Greiner, 2012; Masclet, Noussair and Villeval, 2013; Solda and Villeval, 2017; Ambrus and Greiner, 2019). However, Gächter, Renner and Sefton (2008) document that peer punishment can lead to lower efficiency in the short run than not having any punishment at all. Individuals also seem to favor centralized punishment (or sanctioning) institutions if it is not possible to agree on a norm on socially appropriate behavior (Fehr and Williams, 2018). We add to this literature in two ways. First, we shed light on individuals' preferences over the size of such centralized institutions. We find that a large majority of subjects in our experiment are willing to implement institutions even if there are other subjects who are not bound by the institution's rules but still benefit from the public good. However, around 20% of subjects do not tolerate even a single free-rider and exhibit a strong preference for a grand institution. Second, we document that these preferences are not affected by group size.

The absence of group-size effects on the individual willingness to implement institutions in the presence of single (or few) free-riders also has implications for the literature on how many so-called "bad apples" are tolerated in society (de Oliveira, Croson and Eckel, 2015; Guido, Robbett and Romaniuc, 2019) and the influence of group size on social preferences (Schumacher et al., 2017; Mollerstrom, Strulov-Shlain and Taubinsky, 2021; Alós-Ferrer, García-Segarra and Ritschel, 2022). Our findings indicate that the share of tolerated bad apples (i.e., players who do not want to commit themselves) is not sensitive to small increases in group size. Schumacher et al. (2017) and Alós-Ferrer, García-Segarra and Ritschel (2022) document that social preferences are less influential when decisions affect larger groups. While the group size we study in our experiment is smaller than in either of the two studies, we contribute to this literature by showing that groups with as many as twelve players are not sufficient for social preferences to have no influence on behavior anymore. This is in line with the finding in Mollerstrom, Strulov-Shlain and Taubinsky (2021) that preferences on group-wide redistribution do not seem to depend on group size.

Finally, our findings are related to the more general literature on group-size effects in laboratory experiments. While the group size in our experiment is smaller compared to prior research (e.g., Isaac, Walker and Williams, 1994; Choi, Goyal and Moisan, 2020), we shed light on group-size effects in a previously not-studied domain, namely institution formation games.

The paper proceeds as follows. We describe the institution formation game and outline how money-maximizing and social preferences can affect behavior depending on group size in Section 2. In Section 3 we explain the experimental design and procedures. Section 4 contains our results and we conclude in Section 5.

2 Institution Formation Game

We implement a modified version of the n-person institution formation game analyzed in Okada (1993) and Kosfeld, Okada and Riedl (2009). As part of an institution, players commit themselves to contributing their full endowment to a linear public good, while players who do not commit benefit from other players' contributions and vice versa. Commitments are enforced by assumption, which allows us to focus on the process of institution formation. This setting has been shown to be equivalent to a sanctioning institution which enforces commitments via centralized sanctions (Kosfeld, Okada and Riedl, 2009).[2] The number of players n is common knowledge. The institution formation game consists of the following three stages:

Participation stage: All n players simultaneously and independently indicate their willingness to participate in an institution in which they commit all of their endowment to the public good. We refer to players who indicate their willingness as *participants* and to those who do not as *nonparticipants*. If there is at least one participant, the outcome of the participation stage is an *initiated* institution.

Implementation stage: All players are informed about the number of participants. Participants simultaneously and independently decide whether to accept or reject the implementation of the initiated institution. Only if all participants vote in favor of the implementation (unanimity rule), the institution is successfully implemented and the commitments

[2] Also see McEvoy et al. (2011) who study different enforcement mechanisms in a similar set-up.

become binding. In this case, participants become so-called *members*, while nonparticipants are not part of the institution (so-called *nonmembers*). The implementation of the institution is costly and comes at a fixed cost to each member.[3] If at least a single participant votes against the implementation, the institution is not implemented.[4]

Contribution stage: All players are informed about the implementation outcome of the previous stage and simultaneously and independently decide on their contribution to the public good. If an institution has been implemented, only nonmembers can freely decide on their contribution. Members' commitments are enforced and they contribute their full endowment in this case. If no institution has been implemented, all players freely decide on their contribution.

Let us introduce some further notation. A player's payoff u_i depends on whether an institution is successfully implemented and if yes, her role in it. Each player i has endowment w and we refer to the player's contribution with g_i. The marginal per capita return (MPCR) from contributing to the public good is δ/n with $0 < \delta/n < 1 < \delta$. We thereby hold the social, i.e., total, welfare effect of contributions to the public good constant relative to the group size n. While it is socially efficient to contribute w, individual incentives are such that each player has a strictly dominant action to contribute zero in case no institution is implemented. Let S denote the set of members of the institution with $s = |S|$, and let $c > 0$ be the institution formation cost for each member.

If $S \neq \emptyset$, i.e., an institution is implemented,

$$u_i = \begin{cases} \frac{\delta}{n}\left(sw + \sum_{j \notin S} g_j\right) - c & \text{if } i \in S, \\ w - g_i + \frac{\delta}{n}\left(sw + \sum_{j \notin S} g_j\right) & \text{if } i \notin S. \end{cases} \tag{1}$$

If $S = \emptyset$, i.e., no institution is implemented,

$$u_i = w - g_i + \frac{\delta}{n}\sum_{j=1}^{n} g_j. \tag{2}$$

Equations (1) and (2) highlight that nonmembers of an institution benefit from members' contributions and do not have to bear the institution formation cost c. In addition,

[3]From an applied perspective, this cost may capture, e.g., resources needed to enforce commitments. For our analysis here, the cost is actually irrelevant as long as it not too large (which we ensure both in the model and in the experiment).

[4]If there are no participants, there is naturally no institution to vote upon.

they are free in their contribution decisions. This allows them to free-ride on members' commitments, leading to payoff inequality between members and nonmembers. Finally, we assume that $\delta w - c > w$, i.e., it is socially optimal to form the grand institution with $s = n$.[5]

Before we proceed to the experiment, we outline the role of group size in this set-up. We provide a full equilibrium analysis based on money-maximizing as well as social preferences, using the model of Fehr and Schmidt (1999), in Appendix E. Previous research has highlighted the importance of monetary and fairness concerns in the institution formation process (Kosfeld, Okada and Riedl, 2009; Dannenberg and Gallier, 2020). Let us first consider monetary concerns, i.e., assume money-maximizing preferences. We call an institution *profitable* if members' payoff in an institution is higher than without any institution. If an institution of size s is implemented, members contribute w and nonmembers contribute nothing due to $\delta/n < 1$. If no institution is implemented, nobody contributes and everybody earns w. An institution of size s is thus profitable for a member if

$$\frac{\delta}{n} sw - c > w. \tag{3}$$

Let s^* denote the smallest integer s satisfying this trade-off, characterizing the smallest institution which is profitable to implement. For a fixed δ (as well as w and c), the *relative* size of an institution s/n determines whether an institution is profitable to implement. Moreover, also the relative size of the smallest profitable institution, i.e., s^*/n, is independent of group size.[6] In Appendix E, we show that an institution of size s^*/n is the unique strict equilibrium with an institution implemented on the equilibrium path.[7] We should thus not observe any group-size effects if players were only motivated by monetary concerns.

Next, consider social preferences, i.e., assume that a player's utility depends on the payoff inequality between members (M) and nonmembers (NM). The inequality in payoffs

[5]Of course, the first-best outcome would be if no institution was formed and all players contributed w. However, this outcome is no equilibrium.

[6]Up to rounding. In our setting, we choose the parameters such that the (rounded) s^* is proportional to the increase in n.

[7]Naturally, there always exists an equilibrium in which no institution is implemented and everybody contributes w.

is determined by equation (1) for $s < n$:[8]

$$u_{NM} - u_M = w + c > 0 \tag{4}$$

Notice that the payoff inequality does not depend on n. Let α_i denote a player's aversion to payoff inequality. If players are sufficiently averse to payoff inequality, all institutions other than the grand institution (with $s = n$) are not implemented. In Appendix E, we characterize the threshold $\tilde{\alpha}$ for this to hold. We show that two conditions need to be fulfilled for the grand institution to be the unique equilibrium with an institution implemented on the equilibrium path: there need to be (1) sufficiently many players with (2) sufficiently strong social preferences (i.e., $\alpha_i > \tilde{\alpha}$). If these conditions are fulfilled, social preferences provide an incentive for nonmembers to join, as they do not get the possibility to free-ride on other's contributions.

How does group size affect whether non-grand institutions are implemented in the presence of social preferences? First, social preferences might be sensitive to group size (Schumacher et al., 2017; Alós-Ferrer, García-Segarra and Ritschel, 2022), i.e., individuals might object less to the *same* inequality in larger groups.[9] This would translate into lower α_i. We call this the "preference effect". Second, in Appendix E, we show that the threshold itself, $\tilde{\alpha}$, increases with n. This reflects that increasing group size allows for relatively larger non-grand institutions (e.g., comprising 11 out of 12 compared to 3 out of 4 players) which are more profitable and generate less inequality. We call this the "threshold effect". Both the preference effect and the threshold effect imply that social preferences, while continuing to favor the grand institution, need to be increasingly strong for players to reject non-grand institutions in larger groups.

However, group size mechanically increases the likelihood that there will be "sufficiently many" players with $\alpha_i > \tilde{\alpha}$ in a group. We call this the "composition effect". The overall effect of group size on the implementation of non-grand institutions thus depends on the combination of all three effects. That is, the magnitude of the composition effect and whether the increase in the threshold changes implementation decisions depend on the

[8] For simplicity, we assume here that nonmembers do not contribute to the public good. We relax this assumption in our formal treatment in Appendix E. If (some) nonmembers also contribute to the public good, equation (4) gives the inequality in payoffs between a member and a non-contributing nonmember, while the inequality between a member and a fully contributing nonmember is given by c.

[9] While social preferences might also become stronger with group size, the evidence cited above points to social preferences playing a weaker role in larger groups – albeit in different strategic settings.

initial distribution of social preferences *and* how this distribution is affected by group size (i.e., the preference effect).

Summing up, if players had only money-maximizing preferences, group size should not affect the institution formation process and groups should implement institutions of the same relative size, i.e., s^*/n. Any finding other than that implies the presence of social preferences in small and/or large groups. Notice that this set-up then allows us to disentangle the influence of each of the three effects spelt out above and to infer the role social preferences play at different group sizes. Because institutions of the same *relative* size feature the same profitability and inequality, we can compare implementation decisions across group sizes while holding constant both profitability and inequality. If social preferences, for example, played a much weaker role in larger groups (the preference effect), we should observe that individual players are more likely to accept profitable non-grand institutions of the same relative size. I.e., non-grand institutions should be implemented more frequently.

3 Experimental Design and Procedures

Subjects in our experiment played the institution formation game for 20 rounds in fixed groups (partner matching). At the end, one round was randomly chosen to be payoff-relevant for each group. We adopted a neutral framing for the instructions, following Kosfeld, Okada and Riedl (2009). Before subjects could start with the experiment, they had to answer a set of control questions. See Appendix A for details.

In each round, subjects received 10 points as endowment and the institution costs amounted to 1 point, i.e., $w = 10$ and $c = 1$. In our first treatment, we implemented the same group size $n = 4$ as in Kosfeld, Okada and Riedl (2009); in the other treatments, we doubled and tripled this variable. Every subject only participated in one treatment (between-subject design). This gives us three treatment groups which only differ in the number of players per group, i.e., $n = 4$ (*Treat-S*), $n = 8$ (*Treat-M*), and $n = 12$ (*Treat-L*). We calibrated $\delta = 2.4$ such that the smallest institution size which is profitable to implement corresponds to $s^* = 2, 4, 6$ respectively for the three treatments. The parameter δ was chosen such that our *Treat-S* treatment closely matches the high profitability treatment in Kosfeld, Okada and Riedl (2009). The relative size of the smallest institution which is profitable to implement is 0.5 for all treatment groups.[10]

[10] As we show in Appendix E, this institution size is also the unique strict equilibrium with an institution

We piloted the experiment in August 2020 and conducted the sessions in August and September 2020 with oTree (Chen, Schonger and Wickens, 2016). IRB approval was obtained from the joint ethics committee of Goethe University and the University of Mainz prior to conducting the experiment. Subjects were recruited using ORSEE (Greiner, 2015) from the subject pool of the Frankfurt Laboratory for Experimental Economic Research (FLEX) at Goethe University. Due to COVID-19 social distancing rules at that time, the experiment took place online. In Appendix A, we outline the various protocols we implemented to deal with the online implementation.[11] In total, 308 subjects completed the experiment.[12] 56 subjects participated in treatment *Treat-S* (14 groups), while 120 and 132 participated in treatments *Treat-M* and *Treat-L* (leading to 15 and 11 groups), respectively.[13] On average, the experiment lasted 64 minutes and subjects earned 11.63€ with every point in the experiment worth 0.60€.[14] We do not use the data from the pilot sessions in our analysis.

4 Results

Before looking at the effect of group size on institution formation, we emphasize that *Treat-S* by and large replicates the results from Kosfeld, Okada and Riedl (2009), despite several differences in the parametrization, subject sample, and decision-making environment (online vs. in-person).[15]

In the following, we focus on how group size affects the implementation of profitable

implemented on the equilibrium path assuming money-maximizing preferences.

[11]Snowberg and Yariv (2021) and Prissé and Jorrat (2021) find that subjects recruited from the same subject pool exhibit largely similar behavior in various experimental games regardless of whether the experiment took place online or in the lab.

[12]Appendix A also contains information on how we dealt with dropouts. In short, dropouts occurred rarely and only before the first round had started, e.g., during control questions. Groups with dropouts were excluded from the rest of the experiment, and subjects from these groups were compensated for their time.

[13]Before the institution formation game, subjects also participated in another, unrelated task. However, there was no interaction among subjects in this task and subjects were informed about the outcome only after the overall experiment had ended.

[14]We observe some treatment differences in how long subjects took to complete the experiment with longer times occurring in larger groups on average. In principle, this could lead to negative effects on the willingness to participate in or to implement institutions. Below we show that the willingness to implement an institution is not affected by group size. Additional results (available upon request) show that longer wait times specifically do not affect the willingness to participate in and implement institutions.

[15]Detailed results available upon request.

non-grand institutions.[16] We first analyze whether group size affects these dynamics at all and document a substantial negative effect (Result 1). We then take a closer look at individual implementation decisions to pin down the channel for this effect and infer the role social preferences play in larger groups (Result 2). Finally, we analyze how and why this matters for overall efficiency (Result 3). In order to make the results comparable across treatments, we always report the share of participants (i.e., the relative size of an initiated institution) and members (i.e., the relative size of an implemented institution) in the group instead of the respective absolute numbers.

We start by analyzing whether *aggregate* implementation rates of profitable non-grand institutions are affected by the increase in group size. Figure 1 plots the average group implementation decision, i.e., likelihood that an institution is implemented, conditional on the size of the initiated institution. For convenience, we also illustrate the payoff-optimal voting decision based on the profitability of an institution, i.e., threshold s^* (red line).

Figure 1 clearly highlights that groups have a preference for grand institutions, i.e., the likelihood that a profitable institution which involves less than n players is rejected is quite high.[17] Further, there is a pronounced group-size effect on implementation rates of such institutions. For example, when $n-1$ players participate, we observe an average implementation rate of 0.44 in *Treat-S* vs. 0.12 in *Treat-M* ($p = 0.034$) and 0.28 in *Treat-L* ($p = 0.057$).[18] When comparing implementation rates for institutions with 75% of subjects in a group participating, these differences are even larger: 0.10 in *Treat-M* ($p = 0.007$) and 0.06 in *Treat-L* ($p = 0.008$). Importantly, there is no significant difference when we look at smaller profitable institutions.[19] I.e., group size seems to matter most for the implementation of so-called *almost-grand institution* with at least 75% of the players participating (or put differently, with some but not too much inequality).

Summing up, Figure 1 shows that the emphasis on grand institutions does not only persist but actually gets stronger in larger groups. Almost-grand institutions which are

[16] In Appendix C, we provide a full overview on how group size affects the outcome of each stage.

[17] In Figure 2 we show that almost all participants always vote in favor of the grand institution, i.e., the fact that the implementation rate for grand institutions is less than one is driven by single individuals who occasionally vote against it. This is mirrored in the finding that any differences are *not* significant.

[18] All statistical tests reported are based on group averages as units of independent observations, while the respective p-values refer to pairwise two-sided Mann-Whitney-U tests with the exact p-values unless noted otherwise.

[19] The differences between *Treat-M* and *Treat-L* are not statistically significant ($p \geq 0.671$), neither are any other differences in implementation rates for other profitable institutions for the same relative size, including the grand institution. The results are robust to pooling the implementation rates of all intermediate non-grand institutions to the next smallest institution size which all treatments have in common.

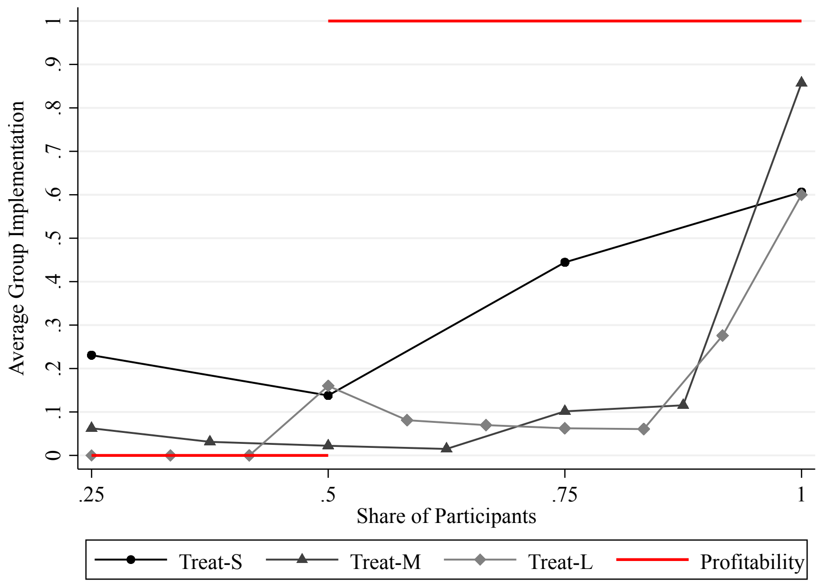


Figure 1: Group Implementation Rates

Notes: Average implementation rates of institutions depending on the share of participants in the group (i.e., the relative size of the initiated institution) across all rounds. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level. The red line indicates which institutions should be implemented based on their profitability alone, i.e., based on s^*.

implemented sometimes in *Treat-S* are almost never implemented in larger groups. Recall that implementation rates for these institutions should not be affected in the case of money-maximizing preferences. In fact, even if social preferences played a role in small but not in large groups, we should observe an increase in the implementation rate of non-grand profitable institutions. This finding therefore indicates that concerns over inequality still matter in larger groups. Result 1 summarizes our findings.

Result 1. *Increasing group size significantly lowers the implementation rate of almost-grand institutions. This indicates that players are not only motivated by monetary concerns in large groups.*

Since this is a complex game and players might learn over time, results might change when we look at later rounds. While we observe that differences in implementation rates become smaller in rounds 11-20, the pattern is qualitatively similar (see Appendix B for details).

Result 1 highlights that the dynamics of institution formation are affected by group size, but it is silent on *why* this is the case. For example, social preferences might indeed play a weaker role in larger groups, but a very strong composition effect could hide this. We therefore turn to individual implementation decisions and focus on almost-grand institutions. Figure 2 replicates Figure 1 on the individual level. It plots participants' average individual voting decision across different group sizes depending on the share of participants, i.e., the relative size of the initiated institution. Recall that our set-up allows

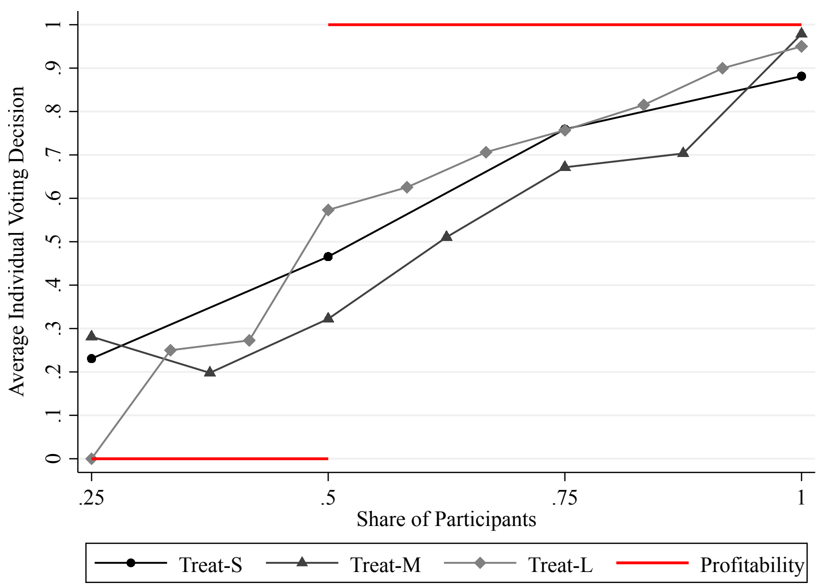


Figure 2: Individual Voting Decisions

Notes: Average individual voting decision depending on the share of participants in the group (i.e., the relative size of the initiated institution) across all rounds. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level. The red line indicates which institutions should be implemented based on their profitability alone, i.e., based on s^*.

us to compare implementation decisions of the same relative institution size across treatments while holding constant profitability and inequality. This helps us in disentangling the different effects and allows to infer the importance of social preferences in larger groups. First, the threshold effect: the preference threshold for rejecting profitable, non-grand institutions increases with group size, because relatively larger almost-grand institutions are possible. E.g., players might tolerate a single free-rider in groups of eight and twelve but not in groups of four. If the threshold effect influenced behavior, i.e., because some players had α_i which are above the threshold for *Treat-S* but below the threshold for *Treat-L*, individuals should vote more frequently in favor of the *largest* non-grand institution in larger groups. While Figure 2 shows that individuals seem to vote more frequently in favor of this institution in *Treat-L* than in *Treat-S*, the opposite is true for *Treat-M*. Moreover, none of these differences is statistically significant ($p \geq 0.290$).

Second, the preference effect: social preferences might play a weaker role in larger groups. Figure 2 indicates that this is not the case, at least on average. Differences between *Treat-S* and any of the two other treatments are not statistically significant when looking at the individual voting decisions for institutions with 75% of players participating.[20] However, the distribution of preferences might have changed such that social preferences play a weaker role for some but a stronger role for other players, leading to no change in the average individual voting decision. Figure 3 therefore plots the distribution of voting decisions on almost-grand institutions, averaged on the level of an individual player, in each treatment.[21] Notice that the majority of subjects in our experiment are consistent in their decision to implement almost-grand institutions, i.e., they always vote in favor of or always vote against them. This is reassuring given our interpretation in terms of preferences. Most participants always vote in favor of the implementation of almost-grand institutions. A smaller yet still substantial share of about 20% of the subjects always vote against the implementation. The shares of these two types are not significantly different across group sizes ($p \geq 0.331$). Further, also the share of subjects whose average willingness to implement almost-grand institutions is at most $x\% \in \{10, 20, 30\}$ or at least $y\% \in \{70, 80, 90, 100\}$ (i.e., these participants voted for the implementation in at most/least $x\%/y\%$ of cases) is not significantly different between *Treat-S* and any of the two larger group size treatments ($p \geq 0.295$). The same holds for the standard deviation in the individual average implementation decisions ($p \geq 0.164$). Thus, group size does not seem to affect the distribution

[20] In fact, no difference between institutions of the same relative size are statistically significant, including the grand institution. Pooling intermediate institution sizes to the next smallest institution all treatments have in common does not lead to a significant difference between *Treat-S* and any of the two larger groups either ($p \geq 0.380$) with the only exception being between *Treat-S* and *Treat-L* for institutions with at least 50% but strictly less than 75% of participants ($p = 0.004$). Comparing only *Treat-M* and *Treat-L*, we find some significant but small differences, with subjects in *Treat-L* exhibiting a higher willingness to implement institutions with at least 50% but strictly less than 75% participants and when pooling all non-grand institutions with at least 75% participants. For *Treat-L*, we observe only a single group that votes on a 25%-institution.

[21] We decided to pool all almost-grand institutions for reasons of statistical power because players' decisions in *Treat-M* and *Treat-L* are spread over a larger set of possible almost-grand institutions.

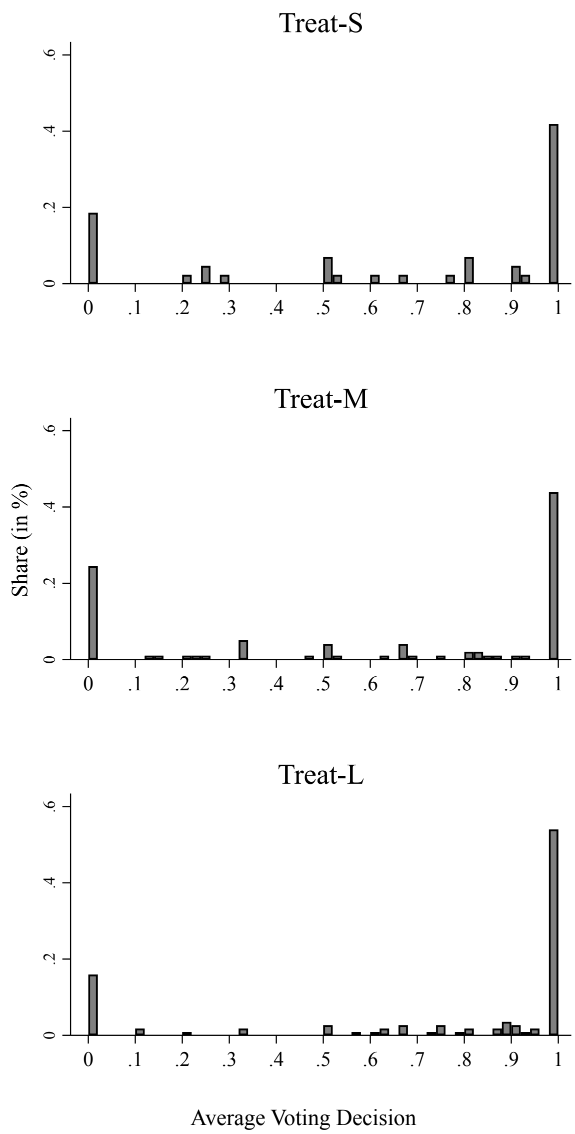


Figure 3: Distribution of Average Voting Decision on Almost-Grand Institutions

Notes: Distribution of the average voting decision on almost-grand institutions (i.e., all non-grand institutions with at least 75% of players participating) across *all* rounds. The average is computed as the average voting decision for each player who participated at least once in an almost-grand institution. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level.

of individual preferences. Moreover, social preferences do not play a weaker role in larger groups.

These findings suggest that the decrease in the implementation rate of almost-grand institutions is exclusively driven by the composition effect: group size mechanically increases the probability of having at least one player per group with sufficiently strong social preferences to block an almost-grand institution. We therefore look at the share of groups with at least one player who always votes against the implementation of almost-grand institutions. This share should increase with group size. This is indeed what we find. The share of groups with at least one such subject increases from 42% in *Treat-S* to 77% in *Treat-M* and even 90% in *Treat-L*. The differences between *Treat-S* and *Treat-M* (p = 0.082, Fisher's one-sided exact test) as well as *Treat-L* are statistically significant (p = 0.026, Fisher's one-sided exact test), while the difference between *Treat-M* and *Treat-L* is statistically insignificant ($p \geq 0.404$). Moreover, the variation in these shares explains the variation in implementation rates of almost-grand institutions both across treatments (Spearman's ρ = -0.7485 with p = 0.000) and within treatments (Spearman's ρ = -0.6242 with p = 0.030 for *Treat-S*, ρ = -0.5609 with p = 0.046 for *Treat-M*, and ρ = -0.7454 with p = 0.013 for *Treat-L*). The same mechanism also explains why we do not find a negative effect of group size on implementation rates of even smaller non-grand yet still profitable institutions. If we look at the share of groups who have at least one subject who always votes against non-grand yet still profitable institutions with less than 75% participants, 91% of groups in *Treat-S* already have at least one such subject. The share increases only a bit for larger group sizes with no statistically significant differences across treatments ($p \geq 0.500$). Our results are robust to looking at subjects who almost always vote against it (instead of always) and looking only at rounds 11-20 (cf. Appendix B). Result 2 summarizes our finding.

Result 2. *The negative effect of group size on the implementation rate of almost-grand institution is exclusively driven by the composition effect. Moreover, social preferences do not play a weaker role in larger groups.*

So far, we have interpreted the decision to vote against the implementation of an almost-grand institutions as evidence of social preferences. To corroborate that interpretation, we perform a simple benchmarking exercise in Appendix C. We calculate the α_i based on the model by Fehr and Schmidt (1999) such that an individual would prefer to reject a non-grand institution comprising at least 75% of group members and compare the share of subjects exhibiting that behavior with the distribution of α_i estimated in Blanco, Engelmann and Normann (2011). In Blanco, Engelmann and Normann (2011), roughly 26.39% of subjects have sufficiently strong preferences to reject almost-grand institutions in our experiment. This result (26.39%) is very comparable to the share of players who always vote against almost-grand institutions in our experiment, i.e., 18.60% in *Treat-S*, 24.49% in *Treat-M*, and 15.93% in *Treat-L*.[22]

We now turn to how overall efficiency is affected by the decrease in the likelihood that institutions involving inequality are implemented. Based on (Kosfeld, Okada and Riedl, 2009), the lower implementation rate of almost-grand institutions in larger groups should, in principle, lead to more grand institutions and thus higher efficiency, as nonparticipants should learn that they have no possibility to free ride on other's contributions. However, Figure 4 shows that this is not the case. The share of grand institutions initiated in stage one decreases from about 37% in *Treat-S* to 12% in *Treat-M* ($p = 0.011$) and even 2% in *Treat-L* ($p = 0.002$). The difference between *Treat-M* and *Treat-L* is not statistically significant ($p = 0.426$). At the same time, more non-grand yet profitable institutions are initiated, for which the share increases from about 49% in *Treat-S* to 70% in *Treat-M* and even 90% in *Treat-L*. All differences are statistically significant ($p \leq 0.057$) and are robust to looking only at rounds 11-20 (cf. Appendix B).

Where does this effect come from? On the individual level, the data suggest that participants very often continue to participate in the next round when an almost-grand institution was not implemented, regardless of group size. This implies that for the grand institution to be initiated, it really is nonparticipants who need to join. The larger group size (and thus the larger number of nonparticipants), however, might make it more difficult in "coordinating" when to join.[23] While this is unlikely – recall that participation decisions are not binding, i.e., nonparticipants have nothing to lose by participating – we still present evidence to shed doubt on miscoordination as a potential reason. If group size was just mechanically lowering the likelihood that a grand institution is initiated through miscoordination, nonparticipants should still join at the same rate in large as in small groups on average. This, however, is not the case. When an almost-grand institution is not implemented, the likelihood of a nonparticipant participating in the following round actually

[22] In Appendix C, we also explore whether subjects who always vote against the implementation of almost-grand institutions are different along observable characteristics (i.e., age, gender, study program, performance in control questions) and do not find any statistically significant differences.

[23] Importantly, note that there is no problem with multiple institutional equilibria in the experiment. As Figure 3 highlights, participants are very consistent in their decision to reject almost-grand institutions. The only equilibrium other than no institution at all therefore is the grand institution.

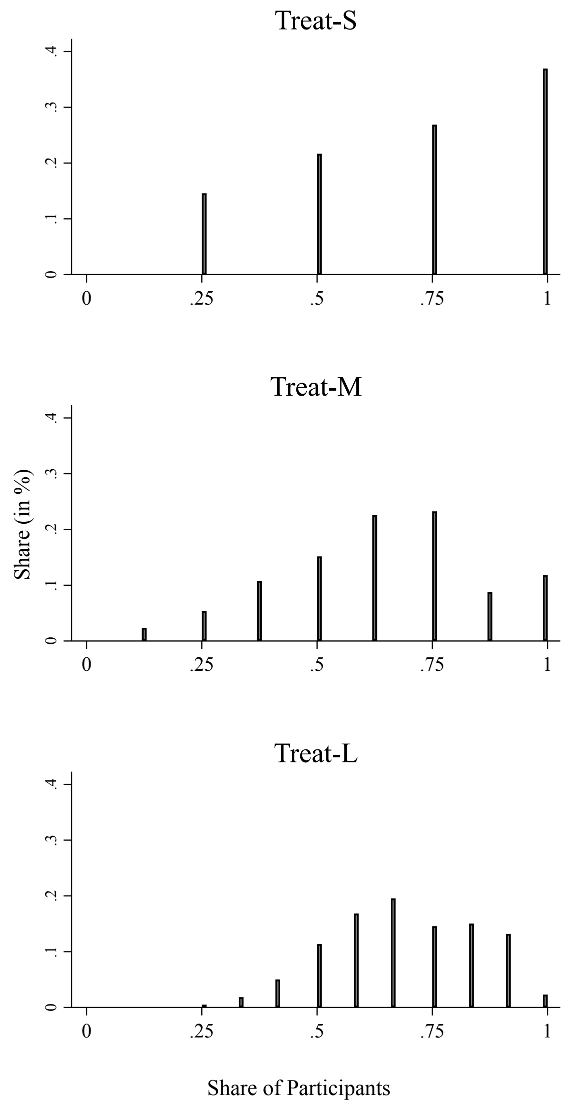


Figure 4: Relative Size of Initiated Institutions

Notes: Distribution of the relative size of the initiated institutions, i.e., the share of participants in the group, over all rounds. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level.

decreases from 45% in *Treat-S* to 28% in *Treat-M* and *Treat-L*. We can also exclude lack of learning among nonparticipants as a potential reason to the extent that the pattern of results is similar when looking at the last 10 rounds (cf. Appendix B). This suggests that nonparticipants either need substantially more periods to learn that almost-grand institutions are rejected or are motivated by other behavioral drivers. We can only speculate here, but conjecture that, for example, social pressure might decline (Olson, 1965) and moral wiggle room (for not committing) might increase (Dana, Weber and Kuang, 2007) as groups become larger.

Since nonparticipants do not join, group size predictably impacts overall implementation rates and efficiency very severely. Panel A in Table 1 displays the overall implementation rate of institutions across treatments, while Panel B shows average contributions to the public good per player and thus overall efficiency. Doubling the group size from four to eight players leads to a pronounced drop in the implementation rates from 41% in *Treat-S* to 15% in *Treat-M* ($p = 0.005$). In the largest groups with twelve players (*Treat-L*), the implementation rate even decreases to 11% ($p = 0.004$). The difference between *Treat-M* and *Treat-L* is not statistically significant ($p = 0.990$).

Table 1: Institution Formation & Efficiency

	Treat-S	*Treat-M*	*Treat-L*
A. Institution Formation			
Implementation Rate	40.67	14.81	11.36
B. Efficiency			
Contributions per Player	5.99	3.17	2.40
If implementation successful	9.16	9.24	8.65
If implementation failed	3.70	2.15	1.60

Notes: Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data refer to averages which are pooled on the treatment level.

Turning to efficiency, while the average contribution per player is quite high in *Treat-S* at 5.99 points (out of 10 points), this almost halves to 3.17 points in *Treat-M* ($p = 0.008$) and more than halves to 2.4 points in *Treat-L* ($p = 0.001$). There is no statistically significant difference between *Treat-M* and *Treat-L* though ($p = 0.609$). Moreover, the decrease in efficiency is mainly driven by the very low implementation rate in larger groups.

While the average contribution per player in case an institution fails to be implemented decreases significantly when group size increases, the absolute differences are relatively small (3.70 in *Treat-S* vs. 2.15 in *Treat-M* with $p = 0.029$ and 1.60 in *Treat-L* with $p = 0.005$), especially compared to the average contribution per player when an institution is successfully implemented. In this case, subjects contribute close to 9 points on average and any difference across treatments is insignificant ($p \geq 0.633$). Result 3 summarizes.

Result 3. *Despite the larger incentive to form grand institutions, fewer grand institutions are initiated in larger groups. Increasing group size therefore reduces the positive impact of institution formation on public good provision significantly.*

5 Conclusion

We study the effect of group size on the impact of commitment institutions for solving collective action problems. Schmidt and Ockenfels (2021) emphasize that such commitment institutions are greatly superior to individual commitments in achieving cooperation. Because decision makers often cannot be forced to cooperate in the first place, participation in such institutions is voluntary and affords decision makers the power to veto commitments. If not everyone joins the institution, commitment institutions generate inequality. Prior research has stressed the positive effect of social preferences on institution formation – albeit in relatively small groups. Existing research also suggests that social preferences might play a weaker role in larger groups, raising the question how group size affects the institution formation process. We moderately increase the group size up to twelve players and find that social preferences influence behavior as strongly in large groups as in small groups. In fact, their influence on group behavior is amplified by the unanimity rule when group size increases beyond four players. Since the lower implementation rate of non-grand institutions does not increase the number of players willing to commit themselves, however, fewer institutions are formed in larger groups. Instead of boosting efficiency, social preferences therefore greatly inhibit the success of institution formation when group size increases.

These results raise the question on how to solve collection action problems through commitment institutions in larger groups. While relinquishing the power to veto in favor of a simple majority rule seems like a solution at first glance, the lack of (existing) institutions to force people to cooperate lies at the heart of many collective action problems. Future research may therefore study in more depth why players do not react to the stronger incentive to form grand institution as well as consider additional instruments which retain the power to veto but are better at reaching a consensus on how much inequality to tolerate, i.e., to harness the power of social preferences. Examples for this from smaller group sizes range from announcing (minimum) institution preferences (Masclet, Noussair and Villeval, 2013; Schmidt and Ockenfels, 2021) and renegotiating if no agreement was reached (He and Villeval, 2017) to communicating before voting (Balafoutas et al., 2014). In the absence of such instruments, our results, however, emphasize that social preferences are a double-edged sword for the ability of commitment institutions to solve collective action problems.

References

Agrawal, Arun, and Gautam Yadama. 1997. "How Do Local Institutions Mediate Market and Population Pressures on Resources? Forest Panchayats in Kumaon, India." *Development and Change*, 28(3): 435–465.

Alós-Ferrer, Carlos, Jaume García-Segarra, and Alexander Ritschel. 2022. "Generous with Individuals and Selfish to the Masses." *Nature Human Behaviour*, 6: 88–96.

Ambrus, Attila, and Ben Greiner. 2012. "Imperfect Public Monitoring with Costly Punishment: An Experimental Study." *American Economic Review*, 102(7): 3317–3332.

Ambrus, Attila, and Ben Greiner. 2019. "Individual, Dictator, and Democratic Punishment in Public Good Games with Perfect and Imperfect Observability." *Journal of Public Economics*, 178: 104053.

Balafoutas, Loukas, Rudolf Kerschbamer, Martin Kocher, and Matthias Sutter. 2014. "Revealed Distributional Preferences: Individuals vs. Teams." *Journal of Economic Behavior & Organization*, 108: 319–330.

Baland, Jean-Marie, and J. P. Platteau. 1996. *Halting Degradation of Natural Resources: Is There a Role for Rural Communities?* Oxford: Oxford University Press.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann. 2011. "A Within-Subject Analysis of Other-Regarding Preferences." *Games and Economic Behavior*, 72(2): 321–338.

Carpenter, Jeffrey P. 2007. "Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods." *Games and Economic Behavior*, 60(1): 31–51.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance*, 9: 88–97.

Cherry, Todd L., Stephan Kroll, and Jason F. Shogren. 2005. "The Impact of Endowment Heterogeneity and Origin on Public Good Contributions: Evidence from the Lab." *Journal of Economic Behavior & Organization*, 57(3): 357–365.

Choi, Syngjoo, Sanjeev Goyal, and Frederic Moisan. 2020. "Large Scale Experiments on Networks A New Platform with Applications." *Cambridge Working Papers in Economics*, 2063.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang. 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory*, 33(1): 67–80.

Dannenberg, Astrid, and Carlo Gallier. 2020. "The Choice of Institutions to Solve Cooperation Problems: A Survey of Experimental Research." *Experimental Economics*, 23(3): 716–749.

Dannenberg, Astrid, Andreas Lange, and Bodo Sturm. 2014. "Participation and Commitment in Voluntary Coalitions to Provide Public Goods." *Economica*, 81(322): 257–275.

D'Aspremont, Claude, Alexis Jacquemin, Jean Jaskold Gabszewicz, and John A. Weymark. 1983. "On the Stability of Collusive Price Leadership." *The Canadian Journal of Economics*, 16(1): 17.

de Oliveira, Angela C. M., Rachel T. A. Croson, and Catherine Eckel. 2015. "One Bad Apple? Heterogeneity and Information in Public Good Provision." *Experimental Economics*, 18(1): 116–135.

Dickinson, David L., David Masclet, and Marie Claire Villeval. 2015. "Norm Enforcement in Social Dilemmas: An Experiment with Police Commissioners." *Journal of Public Economics*, 126: 74–85.

Ertan, Arhan, Talbot Page, and Louis Putterman. 2009. "Who to Punish? Individual Decisions and Majority Rule in Mitigating the Free Rider Problem." *European Economic Review*, 53(5): 495–511.

Fehr, E., and K. M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics*, 114(3): 817–868.

Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90(4): 980–994.

Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature*, 415(6868): 137–140.

Fehr, Ernst, and Tony Williams. 2018. "Social Norms, Endogenous Sorting and the Culture of Cooperation." *CESifo Working Paper Series*, No. 7003. Available at SSRN: https://ssrn.com/abstract=3198185.

Fischbacher, Urs, and Simon Gächter. 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments." *American Economic Review*, 100(1): 541–556.

Gächter, Simon, Elke Renner, and Martin Sefton. 2008. "The Long-Run Benefits of Punishment." *Science*, 322(5907): 1510–1510.

Gangadharan, Lata, Nikos Nikiforakis, and Marie Claire Villeval. 2017. "Normative Conflict and the Limits of Self-Governance in Heterogeneous Populations." *European Economic Review*, 100: 143–156.

Greiner, Ben. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association*, 1(1): 114–125.

Guido, Andrea, Andrea Robbett, and Rustam Romaniuc. 2019. "Group Formation and Cooperation in Social Dilemmas: A Survey and Meta-Analytic Evidence." *Journal of Economic Behavior & Organization*, 159: 192–209.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach. 2006. "The Competitive Advantage of Sanctioning Institutions." *Science*, 312(5770): 108–111.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach. 2014. "On Cooperation in Open Communities." *Journal of Public Economics*, 120: 220–230.

He, Haoran, and Marie Claire Villeval. 2017. "Are Group Members Less Inequality Averse than Individual Decision Makers?" *Journal of Economic Behavior & Organization*, 138: 111–124.

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States.* Vol. 25, Cambridge, Mass.: Harvard University Press.

Hviding, Edvard, and Graham B. K. Baines. 1994. “Community-Based Fisheries Management, Tradition and the Challenges of Development in Marovo, Solomon Islands.” *Development and Change*, 25(1): 13–39.

Isaac, R.Mark, James M. Walker, and Arlington W. Williams. 1994. “Group Size and the Voluntary Provision of Public Goods.” *Journal of Public Economics*, 54(1): 1–36.

Kamei, Kenju, Louis Putterman, and Jean-Robert Tyran. 2015. “State or Nature? Endogenous Formal versus Informal Sanctions in the Voluntary Provision of Public Goods.” *Experimental Economics*, 18(1): 38–65.

Kamei, Kenju, Louis Putterman, and Jean-Robert Tyran. 2019. “Civic Engagement as a Second-Order Public Good: The Cooperative Underpinnings of the Accountable State.” *CEPR Discussion Paper*, No. DP13985. Available at SSRN: https://ssrn.com/abstract=3464512.

Kosfeld, Michael, Akira Okada, and Arno Riedl. 2009. “Institution Formation in Public Goods Games.” *The American Economic Review*, 99(4): 1335–1355.

Markussen, T., L. Putterman, and J.-R. Tyran. 2014. “Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes.” *The Review of Economic Studies*, 81(1): 301–324.

Masclet, David, and Marie-Claire Villeval. 2008. “Punishment, Inequality, and Welfare: A Public Good Experiment.” *Social Choice and Welfare*, 31(3): 475–502.

Masclet, David, Charles N. Noussair, and Marie-Claire Villeval. 2013. “Threat And Punishment In Public Good Experiments.” *Economic Inquiry*, 51(2): 1421–1441.

McEvoy, D. M., J. J. Murphy, J. M. Spraggon, and J. K. Stranlund. 2011. “The Problem of Maintaining Compliance within Stable Coalitions: Experimental Evidence.” *Oxford Economic Papers*, 63(3): 475–498.

Mollerstrom, Johanna, Avner Strulov-Shlain, and Dmitry Taubinsky. 2021. “Preferences for Giving Versus Preferences for Redistribution.” *NBER Working Paper*, No. w29375. Available at SSRN: https://ssrn.com/abstract=3944423.

Okada, Akira. 1993. “The Possibility of Cooperation in an N-Person Prisoners’ Dilemma with Institutional Arrangements.” *Public Choice*, 77(3): 629–656.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups. Harvard Economic Studies.* 21. printing ed., Cambridge, Mass.: Harvard Univ. Press.

Ostrom, Elinor. 1999. "Coping With Tragedies Of The Commons." *Annual Review of Political Science*, 2(1): 493–535.

Page, Talbot, Louis Putterman, and Bulent Unel. 2005. "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency." *The Economic Journal*, 115(506): 1032–1053.

Poteete, Amy R., and Elinor Ostrom. 2004. "Heterogeneity, Group Size and Collective Action: The Role of Institutions in Forest Management." *Development and Change*, 35(3): 435–461.

Prissé, Benjamin, and Diego Jorrat. 2021. "Lack of Control: An Experiment." 21. Available at https://ssrn.com/abstract=3932079.

Rustagi, Devesh. 2022. "The Interdependence of Rules and Civic Values in Commons Management." Available at author's webpage (https://sites.google.com/site/deveshrustagi).

Schmidt, Klaus M., and Axel Ockenfels. 2021. "Focusing Climate Negotiations on a Uniform Common Commitment Can Promote Cooperation." *Proceedings of the National Academy of Sciences*, 118(11): e2013070118.

Schumacher, Heiner, Iris Kesternich, Michael Kosfeld, and Joachim Winter. 2017. "One, Two, Many—Insensitivity to Group Size in Games with Concentrated Benefits and Dispersed Costs." *The Review of Economic Studies*, 84(3): 1346–1377.

Snowberg, Erik, and Leeat Yariv. 2021. "Testing the Waters: Behavior across Participant Pools." *American Economic Review*, 111(2): 687–719.

Solda, Alice, and Marie Claire Villeval. 2017. "Exclusion and Reintegration in a Social Dilemma." *Economic Inquiry*, 58(1): 120–149.

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

Tang, Shui Yan. 1992. *Institutions and Collective Action: Self-Governance in Irrigation.* ICS Press.

A Instructions & Procedures

A.1 Online Implementation Details

All experimental sessions were embedded into a Zoom-meeting with both camera and microphone turned off to protect subjects' privacy/identity. Subjects could only chat with the experimenter and not with each other. We sent out individualized tokens (6 random characters, either uppercase letters or digits) by email 24 hours before a session started. Subjects were only allowed into the meeting with their names set to the individualized token. After all subjects had logged into the meeting, the experimenter gave a short verbal introduction and shared details on how to reach the experimenter in case one had problems with the internet connection (see below). The experimenter then shared a link to the session on oTree and subjects had to enter their individualized token again. Subjects never got to know the tokens of others in their group. Subjects were presented with the instructions and had to prove their understanding of the game by answering a similar set of control questions as in Kosfeld, Okada and Riedl (2009). If subjects answered a question incorrectly three times, they were informed about the right answer and could contact the experimenter in case of questions about why their answers were wrong (most did so). Moreover, subjects were also given time and opportunity to better understand the payoff consequences of different member/nonmember-constellations and average contribution by nonmembers, i.e., to learn which institutions are profitable to implement. For this, subjects could simulate the number of (non)members and the average contribution of nonmembers and were presented with the associated earnings of a member and nonmember on a separate "simulation"-page before the game began. Subjects were also provided with a summary of the relevant instructions underneath each decision. Subjects were paid via PayPal immediately after the session had ended and had agreed to this form of payment when signing up for the session.

Besides the possibility to message the experimenter on Zoom, subjects also received a mobile phone number which they could call, text, or message on Whatsapp in case of questions or problems of any sort (also after the session had ended in case of questions regarding the payment).

A.2 Dealing with Dropouts

In online experiments, subjects can easily drop out from the experiment by just closing the browser window (compared to having to leave the room for normal laboratory experiments). However, subjects can also drop out by just navigating to other websites (in another browser tab) or leave the experiment running in the background while working on other tasks. In order to identify all possible sources of dropouts, we implemented a time limit for each page set to 10 minutes. If a timeout was triggered, a subject was labeled as a dropout. One minute before the timeout, subjects were alerted with a prominent banner that their time to interact with the page was about to run out and the timer was prominently displayed in red in the upper right corner of the website (otherwise it was hidden).

Dropouts were not paid and received a warning that in case of another dropout in a future experiment they would be barred from participating in any future experiment at FLEX. Subjects from the dropout's group were paid depending on how far the experiment had proceeded. If the dropout happened before the end of the second round (or before the first round had even started), subjects received a flat payment of 5 points, since it was not possible to randomly choose a round. If the dropout happened after the second round was successfully completed, a random round was chosen for payout. All subjects were informed about this procedure both verbally in the Zoom session and again as part of the written instructions. All dropouts in the experiment happened before the first round had started.

A.3 Instructions and Screenshots

The instructions are based on the instructions in Kosfeld et al. (2009). The following pages contain the English translation of the German instructions used in the experiment. For simplicity, we do not reproduce the original formatting which is very similar to the screenshots of the decisions screens.

Welcome to the second part of the experiment!

Welcome to the second part of this experiment. Please remember that you can also earn money in this part of the experiment. How much money you earn depends on your decision and the decision of the other participants in this experiment. At the end of the experiment you will receive the money you have earned via PayPal.

It is not allowed to communicate with other participants in this experiment in any way!

If you have a question, please contact us via the chat functionality in Zoom. We will then answer your question. When you click on the "Next"-button, we will explain the instructions for the second part of the experiment in detail. A summary of the relevant instructions will also be displayed together with every decision you will make in this experiment. In this experiment you will still make *all* decisions on your PC. You will still play with "points" in the second part of the experiment. For every point you will receive 0.60€ at the end of the experiment.

General instructions for the second part

Please read the following instructions for the second part very carefully! We will ask you some questions afterwards to check whether you understood the instructions.

The second part of the experiment consists of several rounds. You are still a member of the **same** group of participants as in the first part. Every group therefore has the **same 4 (8/12) participants** as in the first part of the experiment. But now you play with **all 3 (7/12) other group members together**. The group composition will not change. You will still receive no information regarding the identity of the other group members – also not after the end of the experiment. The other participants in the experiment will also not receive any information on your identity. Any interaction in this experiment is therefore completely anonymous.

The second part of the experiment consists of **20 rounds**. All rounds are identical and consist of **4 phases**. All participants receive **10 points** at the beginning of each round. In the following, we will refer to these 10 points as **endowment**. Every round is structured as follows: in the **fourth** phase of each round, all group members receive information on the decisions which were taken during the previous phases of the round and the resulting earnings. In the **third** phase of each round, each of the 4 (8/12) group members decides how much of the endowment she wants to contribute to a common project and how much she wants to keep for herself. In the first **two phases** of each round, all group members decide whether they want to **commit themselves to contributing all of their endowment to the project**.

One of the 20 rounds will be **randomly selected** to be paid out at the end of experiment. When you will have completed the 20th round, you will be informed about which round was selected. Since you do not know which round will be selected by the random number generator, you should carefully consider your decisions in all rounds.

If a drop-out should happen in the second part of the experiment, you will receive the earnings from the first part of the experiment and a compensation for the second part. The amount of the compensation depends on how many rounds will have been played when the drop-out happens. If the drop-out happens before the end of the second round, it is not possible to randomly select a round. In this case you will receive 5 points as compensation. If the drop-out happens after the end of the second round, a previously played round will be randomly selected.

On the following pages all decisions in each phase of a round will be explained in detail. A summary of the instructions will also be made available every time you make a decision.

Instructions regarding your earnings

We will start by explaining how your earnings in a given round depend on your decisions in the third phase. On the following pages, we will explain the decisions in the first three phases. Please read the instructions very carefully.

In the third phase all 4 (8/12) group members need to decide how to allocate their endowment (10 points) between a personal account and a common project. Every point which you do not contribute to the common project is automatically transferred to your personal account. Your group members make exactly the same decision with the same rules governing their earnings.

Your earnings from the personal account: for every point which is transferred to your personal account, **you** receive 1 point and the **other** group members 0 points.

Your earnings from the common project: for every point which is contributed to the common project, **you** receive 0.6 (0.3/0.15) points and the **other** group members 0.6 (0.3/0.15) points. The sum of all earnings of all group members from the common project therefore increases by 2.4 points for every point you contribute to the common project. Your contribution to the project thus also increases the earnings of the other group members.

Your earnings from others' decisions: **regardless of how much you contribute to the project**, you receive 0.6 (0.3/0.15) points for every point contributed to the common project by another group member.

Potential costs: in addition to the earnings from your personal account and the common project, there can be costs which are deducted from your earnings. On the following page, we explain when these costs apply.

Your total earnings: your total earnings from this decision situation are the sum of the earnings from your personal account and the project, with any potential costs subtracted from it:

Earnings = (10 - your contribution to the project) + 0.6 × the sum of project contributions of **all** 4 group members - potential costs

This decision situation is therefore similar to the first part of the experiment. The second part is different from the first part in that you are not matched with one randomly chosen person. Instead you are playing with all 3 (7/11) other group members simultaneously. Additionally, group members now have the opportunity to engage in commitments. In the following we explain how these commitments (and any potential costs) can matter.

Instructions regarding your decisions

Now we explain the decisions in all three phases.

First phase

In the first phase you decide whether you are **willing to commit yourself to contributing all 10 points** of your endowment to the **common project** in the third phase. Moreover, we ask you to estimate **how many of your 3 (7/11) other** group members are willing to commit themselves to contributing their endowment to the common project. When you are satisfied with your decisions, please click "Next". When every group member has made these decisions, the next phase starts.

Second phase

All group members who **were willing to commit to contributing all of their endowment** in the first phase now need to decide whether they **actually want to implement these commitments**. This happens as follows. If you were willing to commit yourself in phase 1, you now need to decide whether you **still want to commit** to contributing all of your endowment (10 points) to the common project. For this decision, you will receive information on the number of other group members who were willing to commit their endowment to the project in phase 3. Only if **all group members** who were **willing to commit themselves** in phase 1 are still willing to do so, the commitments **are implemented and become binding**. If, however, **at least one** group member – who was willing to commit herself – does not want to commit herself anymore, **all commitment** are **not implemented and do not become binding**.

Important: only group members who were willing to commit themselves in phase 1 make these implementation decisions. But all group members receive information on how many other group members were willing to commit themselves. **All** group members therefore receive **exactly the same** information, regardless of their decisions in the first phase.

If every group member who was willing to commit herself in phase 1 has made her decision, the next phase starts.

Third phase

In the third phase you have to choose your contribution to the project.

If the commitments became binding, the contribution decision will be automatically made for all group members who have committed themselves. These group members automatically contribute all of their endowment (10 points) to the project. Additionally, these group members also need to pay **extra costs**. These costs are 1 point and are paid by every group member who has successfully committed herself.

If the commitment did not become binding, every group member can freely decide how many points of her endowment she wants to contribute to the project. In this case, there are **no** extra costs.

Important: the commitments (if they became binding) only apply to the group members who were willing to commit themselves in phase 1. Moreover, only those group members need to pay the extra costs. Group members who were not willing to commit themselves, can still freely choose their contribution to the project. When all group members have made this decision, the next phase starts.

Fourth phase

In the fourth (and last) phase of each round you receive information on how many points were contributed to the project in total. Moreover, we will show you your earnings of this round.

Please read these instructions very carefully. If you have questions, please write us in the Zoom chat. We will then respond to your question. When you have understood the instructions, you can click on "Next".

Simulation of earnings

[Please see the respective screenshot in Figure A.1.]

Control Questions

[All control questions contained a summary of the instructions and an online calculator.]

You will now see a few questions which we ask you to answer. These questions are meant to ensure that you have understood the instructions. You do not receive any money for answering these questions, but you have to answer them correctly to participate in the experiment. All questions are based on fictive examples. In these questions, we refer to the group members as A, B, C, and D (A, B, C, D, E, F, G, and H; A, B, C, D, E, F, G, H, I, J, K, and L).

Question 1: Suppose that nobody has committed herself to contributing all of her endowment to the project. Moreover, all 4 (8/12) group members contribute **0 points** to the project. What are the earnings of a single player in points?

Question 2: Suppose that nobody has committed herself to contributing all of her endowment to the project. Moreover, all 4 (8/12) group members contribute **10 points** to the project. What are the earnings of a single player in points?

Question 3a: Suppose that nobody has committed herself to contributing all of her endowment to the project. Player A (A and E; A, E, and I) contributes **2 points**, player B (B and F; B, F, and J) contributes **4 points**, player C (C and G; C, G, and K) contributes **5 points**, and player D (D and H; D, H, and L) contributes **9 points** to the common project. What are the earnings of player B in points?
Question 3b: What are the earnings of player D (H; L) in points?

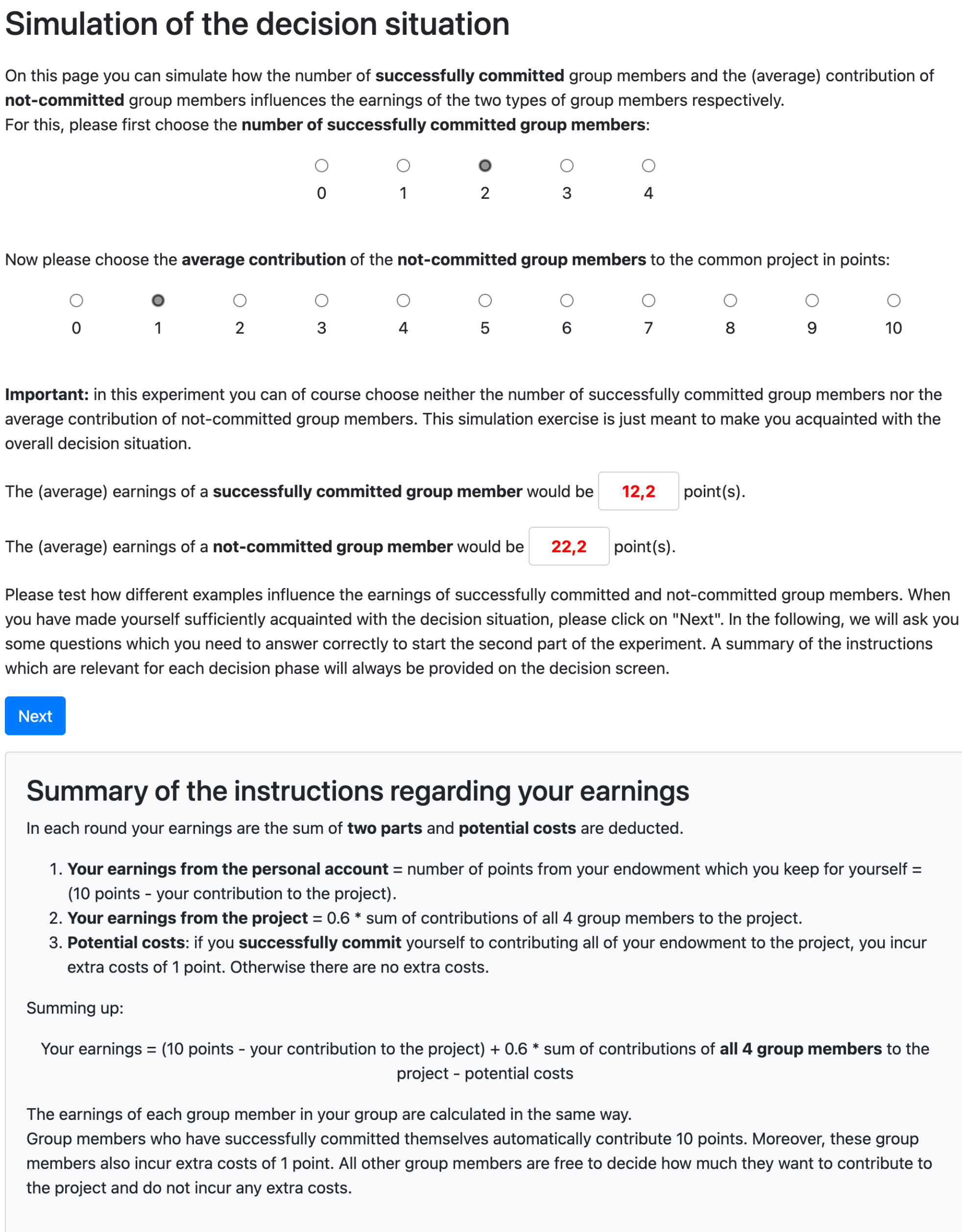
Question 4a: Suppose that in the first phase half of all group members (players A, D; players A, D, E, H; players A, D, E, H, I, L) were willing to commit to contributing all of

their endowment (10 points) to the project. Also suppose that these commitments have become **binding** in phase 2. In phase 3, the other group members (players B, D; players B, D, F, G; players B, D, F, G, J, K) contribute **0 points** points to the common project. How many points does player A (E/I) contribute to the project?
Question 4b: How high are the additional costs for player D in points? **Question 4c**: How high are the additional costs for player B (F/J) in points?

Question 5a: Suppose that in the first phase half of all group members (players A, D; players A, D, E, H; players A, D, E, H, I, L) were willing to commit to contributing all of their endowment (10 points) to the project. Also suppose that these commitments have become **binding** in phase 2. In phase 3, the other group members (players B, D; players B, D, F, G; players B, D, F, G, J, K) contribute **0 points** points to the common project. What are the earnings of player A in points?
Question 5b: What are the earnings of player C (G/K) in points?

Question 6a: Suppose that in the first phase half of all group members (players A, D; players A, D, E, H; players A, D, E, H, I, L) were willing to commit to contributing all of their endowment (10 points) to the project. Also suppose that these commitments have **not become binding** in phase 2. In phase 3, all group members contribute **0 points** points to the common project. How high are the additional costs for player A (E/I) in points?
Question 6b: What are the earnings of player A in points?
Question 6c: What are the earnings of player B (F/J) in points?

[Please see Figures A.2 to A.7 for screenshots of the decision screens.]



Simulation of the decision situation

On this page you can simulate how the number of **successfully committed** group members and the (average) contribution of **not-committed** group members influences the earnings of the two types of group members respectively.
For this, please first choose the **number of successfully committed group members**:

○ 0 ○ 1 ◉ 2 ○ 3 ○ 4

Now please choose the **average contribution** of the **not-committed group members** to the common project in points:

○ 0 ◉ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

Important: in this experiment you can of course choose neither the number of successfully committed group members nor the average contribution of not-committed group members. This simulation exercise is just meant to make you acquainted with the overall decision situation.

The (average) earnings of a **successfully committed group member** would be **12,2** point(s).

The (average) earnings of a **not-committed group member** would be **22,2** point(s).

Please test how different examples influence the earnings of successfully committed and not-committed group members. When you have made yourself sufficiently acquainted with the decision situation, please click on "Next". In the following, we will ask you some questions which you need to answer correctly to start the second part of the experiment. A summary of the instructions which are relevant for each decision phase will always be provided on the decision screen.

Next

Summary of the instructions regarding your earnings

In each round your earnings are the sum of **two parts** and **potential costs** are deducted.

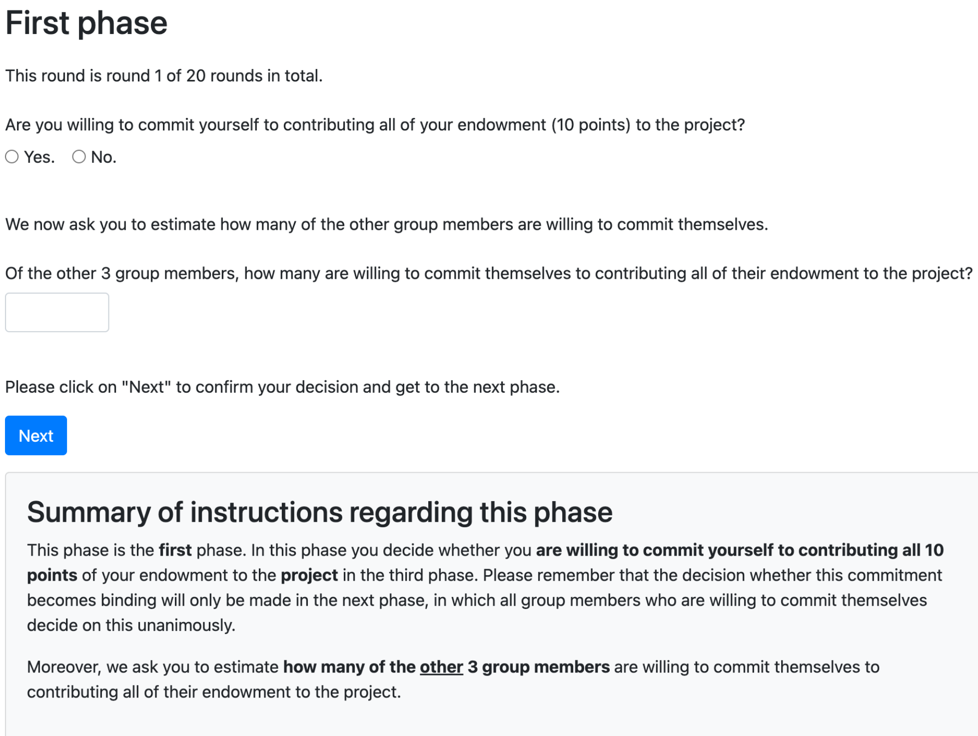
1. **Your earnings from the personal account** = number of points from your endowment which you keep for yourself = (10 points - your contribution to the project).
2. **Your earnings from the project** = 0.6 * sum of contributions of all 4 group members to the project.
3. **Potential costs**: if you **successfully commit** yourself to contributing all of your endowment to the project, you incur extra costs of 1 point. Otherwise there are no extra costs.

Summing up:

Your earnings = (10 points - your contribution to the project) + 0.6 * sum of contributions of **all 4 group members** to the project - potential costs

The earnings of each group member in your group are calculated in the same way.
Group members who have successfully committed themselves automatically contribute 10 points. Moreover, these group members also incur extra costs of 1 point. All other group members are free to decide how much they want to contribute to the project and do not incur any extra costs.

Figure A.1: Simulation Page



First phase

This round is round 1 of 20 rounds in total.

Are you willing to commit yourself to contributing all of your endowment (10 points) to the project?

○ Yes. ○ No.

We now ask you to estimate how many of the other group members are willing to commit themselves.

Of the other 3 group members, how many are willing to commit themselves to contributing all of their endowment to the project?

Please click on "Next" to confirm your decision and get to the next phase.

Next

Summary of instructions regarding this phase

This phase is the **first** phase. In this phase you decide whether you **are willing to commit yourself to contributing all 10 points** of your endowment to the **project** in the third phase. Please remember that the decision whether this commitment becomes binding will only be made in the next phase, in which all group members who are willing to commit themselves decide on this unanimously.

Moreover, we ask you to estimate **how many of the other 3 group members** are willing to commit themselves to contributing all of their endowment to the project.

Figure A.2: First Phase (Participation Phase)

Second phase

This round is round 1 of 20 rounds in total.

Of the **other 3** group members, **2 group members** are willing to commit themselves, **1 group member** is not willing. You are **willing** to commit yourself to contributing your endowment (10 points) to the project.

Do you want to bindingly commit yourself to contributing all of your endowment (10 points) to the project?

○ Yes. ○ No.

Please click on "Next" to confirm your decision and get to the next phase.

Next

Summary of instructions regarding this phase

This phase is the **second** phase. In this phase all group members who, in the first phase, were willing to commit themselves to contributing all of their endowment to the project decide whether to **bindingly** commit themselves.
If **all group members** who were willing to commit themselves in the first phase **are still** willing to commit themselves, these commitments become **binding** and **come into effect**.
If **at least one** group member - who was willing to commit herself - does **not want to commit herself anymore**, these commitment do not come into effect for **any group member** and are thus **not binding**.

Group members who were **not** willing to commit themselves do not make any decision.

Figure A.3: Second Phase for Participants (Implementation Phase)

Second phase

This round is round 1 of 20 rounds in total.

Of the **other 3** group members **all 3 group members** are willing to commit themselves. You are **not willing** to commit yourself to contributing your endowment (10 points) to the project.

Please click "Next" to get to the next phase.

Next

Summary of instructions regarding this phase

This phase is the **second** phase. In this phase all group members who, in the first phase, were willing to commit themselves to contributing all of their endowment to the project decide whether to **bindingly** commit themselves.
If **all group members** who were willing to commit themselves in the first phase **are still** willing to commit themselves, these commitments become **binding** and **come into effect**.
If **at least one** group member - who was willing to commit herself - does **not want to commit herself anymore**, these commitment do not come into effect for **any group member** and are thus **not binding**.

Group members who were **not** willing to commit themselves do not make any decision.

Figure A.4: Second Phase for Nonparticipants (Implementation Phase)

Third phase

This round is round 1 of 20 rounds in total.

The commitments have **become binding** for 3 group members. Your commitment has become binding. You therefore automatically contribute all of your endowment (10 points) to the project.

Please click on "Next" to get to the next phase.

Next

Summary of instructions regarding this phase

This phase is the **third** phase. In this phase all group members decide how much to contribute to the common project.
If the commitments became binding, the contribution decision will be automatically made for all group members who have committed themselves. These group members automatically contribute all of their endowment (10 points) to the project.
Additionally, these group members also need to pay **extra costs**. These costs are 1 point and are paid by every group member who has successfully committed herself.

If the commitment did not become binding, every group member can freely decide how many points of the endowment she wants to contribute to the project. In this case, there are **no** extra costs.

The commitment (if they became binding) only concern the group members who were willing to commit themselves in phase 1. Moreover, only those group members need to pay the extra costs. Group members who have not committed themselves, can still freely choose their contribution to the project.

Your earnings in this rounds are as follows:

Your earnings = (10 points (endowment) - 10 points (your commitment)) + 0.6 * sum of contributions of **all 4 group members** to the project - 1 point (extra costs)

Figure A.5: Third Phase for Members (Contribution Phase)

Third phase

This round is round 1 of 20 rounds in total.

The commitments have **become binding** for 3 group members. You have **not committed** yourself and can freely decide on your contribution to the project. How many points do you want to contribute to the project?

○	○	○	○	○	○	○	○	○	○	○
0	1	2	3	4	5	6	7	8	9	10

Please click on "Next" to confirm your decision and get to the next phase.

Next

Summary of instructions regarding this phase

This phase is the **third** phase. In this phase all group members decide how much to contribute to the common project.
If the commitments became binding, the contribution decision will be automatically made for all group members who have committed themselves. These group members automatically contribute all of their endowment (10 points) to the project.
Additionally, these group members also need to pay **extra costs**. These costs are 1 point and are paid by every group member who has successfully committed herself.

If the commitment did not become binding, every group member can freely decide how many points of the endowment she wants to contribute to the project. In this case, there are **no** extra costs.

The commitment (if they became binding) only concern the group members who were willing to commit themselves in phase 1. Moreover, only those group members need to pay the extra costs. Group members who have not committed themselves, can still freely choose their contribution to the project.

Your earnings in this rounds are as follows:

Your earnings = (10 points (endowment) - your contribution to the project) + 0.6 * sum of contributions of **all 4 group members** to the project

Figure A.6: Third Phase for Nonmembers (Contribution Phase)

Results of round 1 (fourth phase)

Of the **other group members**, 2 group members were willing to commit themselves.

The commitments have **become binding** for 3 group members. You have **committed** yourself. Here are the results:

Your contribution to the project:	10 points
Sum of all contributions:	31 points
Your earnings from the personal account:	0 points
Your earnings from the project:	18.6 points
Extra costs:	1 point
Your earnings in this round:	17.6 points

Please click on "Next" to get to the next phase.

Next

Figure A.7: Fourth Phase (Results)

B Robustness Checks

B.1 Result 1

Result 1 shows that increasing group size significantly decreases implementation rates of almost-grand institutions. This is also true when we account for learning and only consider the last 10 rounds. Figure A.8 replicates Figure 1 for the last 10 rounds. There still exists a marked difference between small and larger groups. Institutions with three out of four subjects are more frequently implemented than institutions with seven (eleven) out of eight (twelve) subjects. By restricting the analysis to a subset of rounds, we do also lose some observations since not every group initiated an almost-grand institution in rounds 11-20. The statistical tests are thus based on fewer groups with the consequence that only the difference between *Treat-S* and *Treat-L* for the 75% institution remains statistically significant ($p = 0.095$) at conventional levels ($p = 0.142$ for the difference in implementation rate of 75% institutions between *Treat-S* and *Treat-M*).[24]

B.2 Result 2

B.2.1 Learning

Result 2 shows that the decrease in the implementation rates of almost-grand institutions is exclusively driven by the composition effect, i.e., the mechanical increase in the likelihood that at least one subject per group has sufficiently strong preferences to block the implementation of almost-grand institutions. This is also true when we account for learning and only consider the last 10 rounds. See Figure A.9 which replicates Figure 2 for rounds 11-20. The threshold effect, i.e., larger group sizes allow for relatively larger almost-grand institutions with resulting lower (average) inequality and higher profitability, would imply that the implementation rates of the largest non-grand institution increases with group size. This is also not the case in the last 10 rounds. Average individual voting decisions are not significantly different between *Treat-S* and any of the two larger group treatments when we consider the largest non-grand institution in each treatment ($p \geq 0.334$), 75% institutions ($p \geq 0.584$), or 50% institutions ($p \geq 0.107$). Regarding the preference effect, Figure A.10 replicates Figure 3 for the last 10 rounds, i.e., rounds 11-20. It clearly shows

[24] To be specific, while the statistical tests for all all rounds are based on 12 groups for *Treat-S* as well as 13 and 9 groups for *Treat-M* and *Treat-L* respectively, the number of groups reduces to 9 (*Treat-S*) and 8 (*Treat-M* and *Treat-L*).

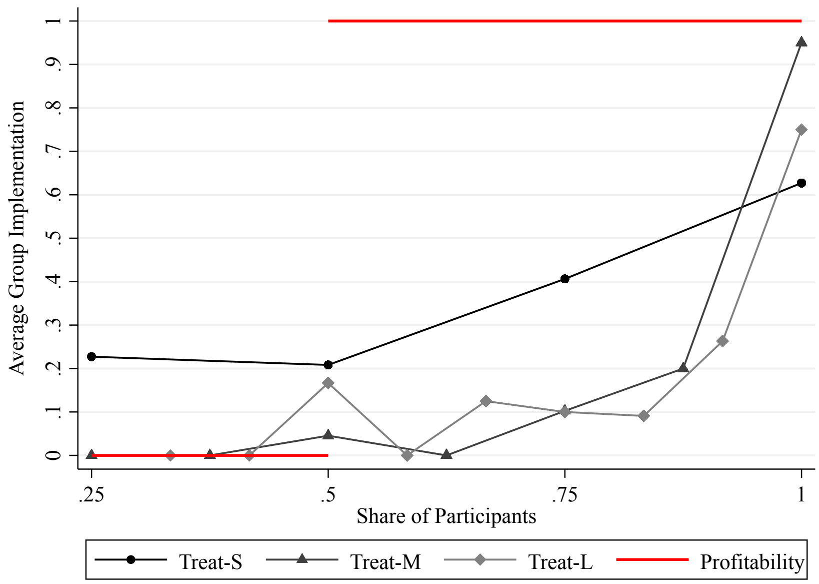


Figure A.8: Group Implementation Rates (Rounds 11-20)

Notes: Average implementation rates of institutions depending on the share of participants in the group (i.e., the relative size of the initiated institution) for the last 10 rounds (rounds 11-20). Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level. The red line marks the non-social prediction according to s^* and indicates which institutions should be implemented based on their profitability alone. The red line indicates which institutions should be implemented based on their profitability alone, i.e., based on s^*. No institution with $s/n = 0.25$ was initiated for *Treat-L*.

that group size does not affect individual implementation decisions over almost-grand institutions. Similar to the main analysis, the great majority of subjects consistently vote in favor of the implementation of almost-grand institutions. Again, roughly 20% of subjects, however, always vote against the implementation of such institutions. Any difference in the share of these subjects is not significantly different across group sizes ($p \geq 0.202$). Moreover, also the share of subjects whose average willingness to implement almost-grand institutions is at most $x \in \{10, 20, 30\}$ or at least $y \in \{70, 80, 90, 100\}$ (i.e., they voted for the implementation in at most/least x%/y% of cases when they participated) is not significantly different between Treat-S and any of the two larger group size treatments ($p \geq 0.202$). Consequently, social preferences themselves do not seem to play a weaker role in larger group sizes.

Finally, we also document for rounds 11-20 that the share of groups with at least one

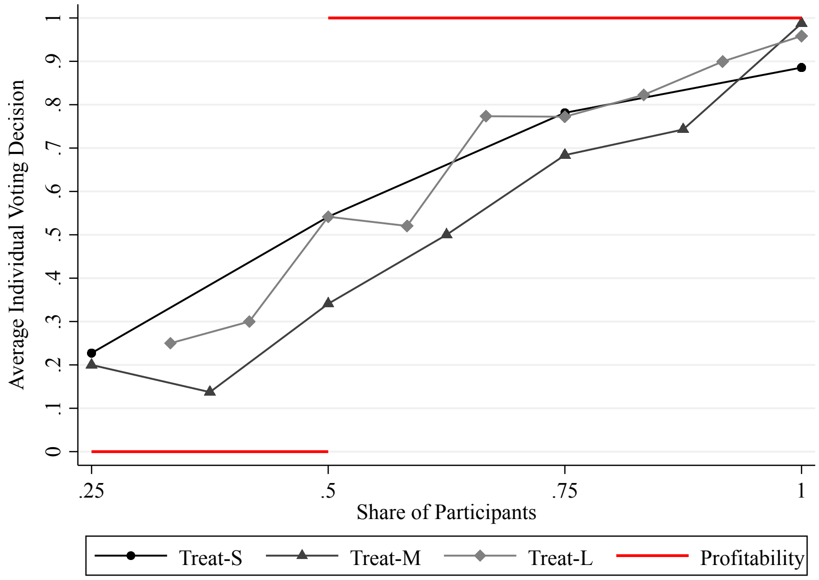


Figure A.9: Individual Voting Decisions (Rounds 11-20)

Notes: Average individual voting decision depending on the share of participants in the group (i.e., the relative size of the initiated institution) for the last 10 rounds (rounds 11-20). Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level. The red line indicates which institutions should be implemented based on their profitability alone, i.e., based on s^*. No institution with $s/n = 0.25$ was initiated for *Treat-L*.

subject who always votes against the implementation of almost-grand institutions strongly increases when increasing group size beyond four players. While only 44.44% of groups have one such player in *Treat-S*, this increases to 88.89% for both *Treat-M* and *Treat-L* ($p =$ 0.066). Similarly, having at least one player who always votes against the implementation of almost-grand institutions also explains the variation in the implementation rates of almost-grand institutions in rounds 11-20 (Spearman's ρ = -0.7771 with p = 0.000). I.e., the composition effect seems to be the main channel through which group size affects the implementation of almost-grand institutions.

B.2.2 Almost-Always Rejecting Almost-Grand Institutions

In our main analysis, we restrict our attention to subjects who *always* vote against the implementation of almost-grand institutions. Our results, however, are also robust to

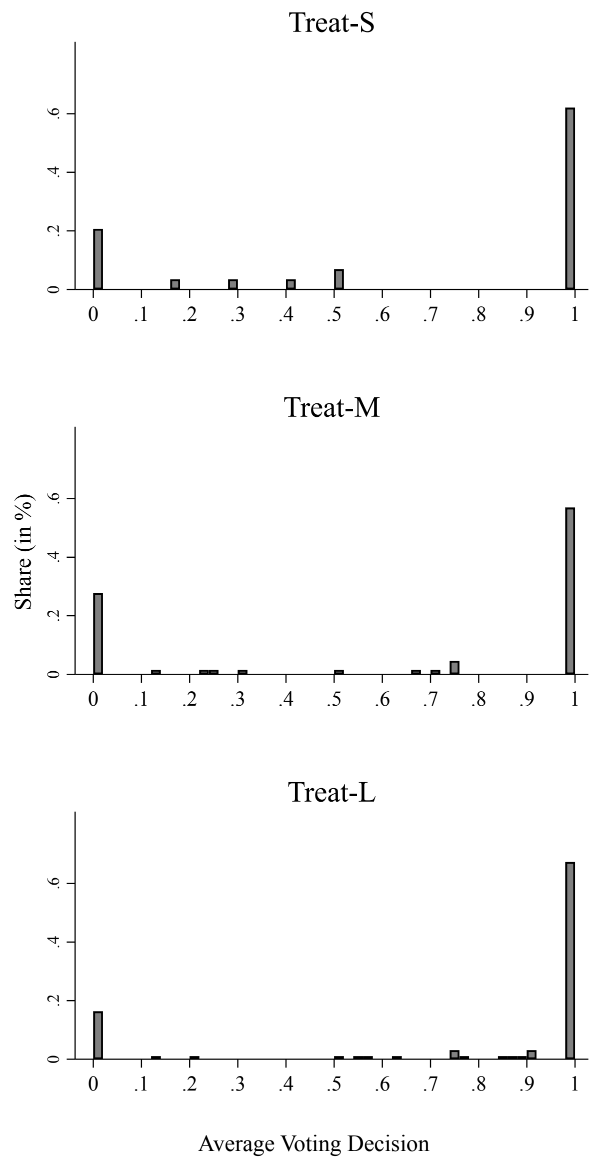


Figure A.10: Distribution of Average Voting Decision on Almost-Grand Institutions (Rounds 11-20)

Notes: Distribution of the average voting decision on almost-grand institutions (i.e., all non-grand institutions with at least 75% of subjects in a group participating) in rounds 11-20. The average is computed as the average voting decision for each player who participated at least once in an almost-grand institution in rounds 11-20. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level.

looking at subjects who *almost always* vote against the implementation of almost-grand institution. The share of subjects who reject almost-grand institutions in at least 80% of times (i.e., whose average willingness to implement almost-grand institutions is at most 20%) is very comparable across group sizes (cf. Figure 3 and not significantly different with $p \geq 0.299$).[25] The share of groups with at least one such subject increases from 41.67% for *Treat-S* to 84.62% for *Treat-M* ($p = 0.033$) and 90.00% for *Treat-L* ($p = 0.026$). This can also explain the variation in implementation rates of almost-grand institutions (Spearman's ρ = -0.7331 with $p = 0.000$).

B.3 Result 3

Result 3 shows that the lower implementation rate of almost-grand institutions in larger groups does not lead to more grand institutions being initiated, which reduces the success of institution formation and public good provision significantly. This finding is again robust to accounting for learning, i.e., only looking at rounds 11-20. In our main analysis, we provided two pieces of evidence for Result 3: the distribution of initiated institutions and the likelihood that nonparticipants change their mind. First consider the distribution of initiated institutions. Figure A.11 replicates Figure 4 for rounds 11-20. Clearly, grand institutions are initiated significantly less frequently in larger groups ($p = 0.012$ for the difference between *Treat-S* and *Treat-M* and $p = 0.006$ for the difference between *Treat-S* and *Treat-L*).

Second, also in rounds 11-20, nonparticipants are less willing to participate in the following round when an almost-grand institution was rejected. The likelihood that a nonparticipant participates in the following round decreases from 47.62% in *Treat-S* to 22.22% in *Treat-M* ($p = 0.065$) and 26.04% in *Treat-L* ($p = 0.040$). Result 3 finds that the effect of group size on the formation dynamics in the presence of inequality has strong economic effects, with overall implementation rates and efficiency decreasing greatly. We also document this effect when accounting for learning, i.e., when looking at rounds 11-20 only. Table A.1 replicates Table 1 for rounds 11-20. Overall implementation rates and average contributions per player are significantly different: implementation rates decrease from 43.80% in *Treat-S* to 17.69% in *Treat-M* ($p = 0.021$) and 14.55% in *Treat-L* ($p = 0.021$). Efficiency also is significantly different ($p = 0.004$ for the difference between *Treat-*

[25] Inspecting Figure 3 shows that there are no additional observations if we set the threshold to 10% instead of 20%.

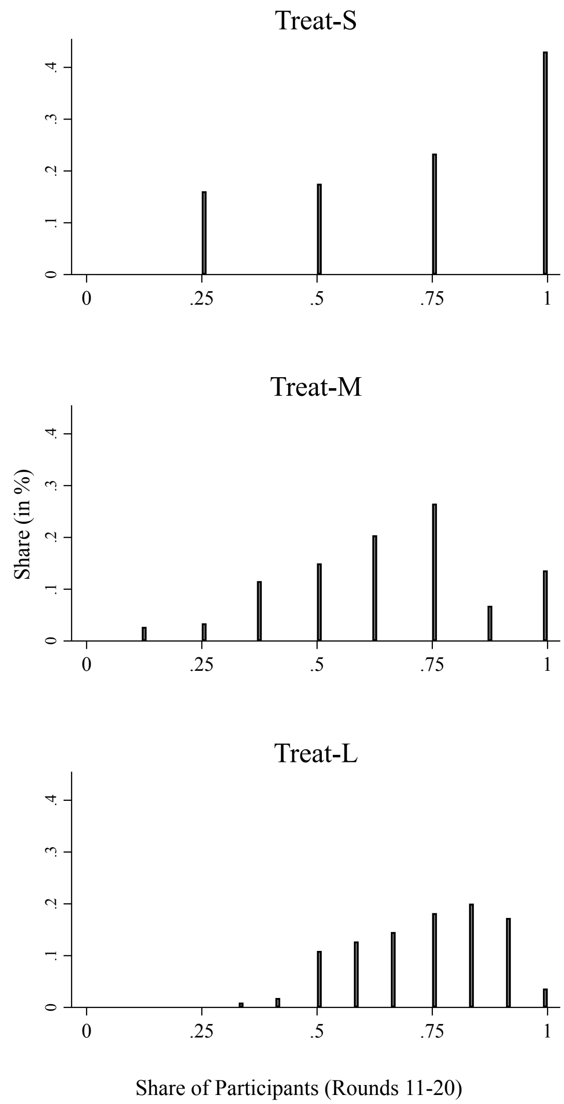


Figure A.11: Relative Size of Initiated Institutions (Rounds 11-20)

Notes: Distribution of the relative size of the initiated institutions, i.e., the share of participants in the group, in Rounds 11-20. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level.

S and *Treat-M* and 0.001 for the difference between *Treat-S* and *Treat-L*) and this reduction is again driven by the decrease in overall implementation rates.

Table A.1: Institution Formation & Efficiency (Rounds 11-20)

	Treat-S	*Treat-M*	*Treat-L*
A. Implementation Success			
Implementation Rate	43.80	17.69	14.55
B. Efficiency			
Contributions per Player	6.10	2.76	2.07
If implementation successful	9.33	9.39	8.74
If implementation failed	3.57	1.39	0.94

Notes: Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data refer to averages which are pooled on the treatment level. Data come from rounds 11-20.

C Additional Analyses

C.1 Average Outcome in Each Stage

Table A.2 lists the average outcome of each formation stage separately for each group size.

Table A.2: Effect of Group Size on Outcomes in Each Stage

	Treat-S	*Treat-M*	*Treat-L*
A. Initiated Institutions			
Number	268	297	220
Frequency (in %)	95.71	99.00	100.00
Participants (avg., in % of group)	71.55	64.23	69.09
B. Implemented Institutions			
Number	109	44	25
Frequency (in %)$^{***/\dagger\dagger\dagger}$	40.67	14.81	11.36
Members (avg., in % of group)	82.80	90.06	77.00
C. Contributions per Player			
Average$^{***/\dagger\dagger\dagger}$	5.99	3.17	2.40
Average (successful implementation)	9.16	9.24	8.65
Average (failed implementation)$^{**/\dagger\dagger\dagger}$	3.70	2.15	1.60

Notes: Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level. $^*/^\dagger/^\diamond$ denote the statistical significance of pairwise two-sided Mann-Whitney-U tests with * (†) referring to the difference between *Treat-S* and *Treat-M* (*Treat-L*) and $^\diamond$ referring to the difference between *Treat-M* and *Treat-L*, based on group-level averages as units of observation.
$^\circ p < 0.1$, $^{\circ\circ} p < 0.05$, $^{\circ\circ\circ} p < 0.01$ (exact p-values).

C.2 Subjects Always Rejecting Almost-Grand Institutions

Table A.3 explores whether subjects who always vote against the implementation of almost-grand institutions (from now on referred to as “nay-sayers”) are different from people who vote at least once in favor of the implementation of an almost-grand institution in terms of their performance in the 12 control questions (total number of incorrect answers), age, sex, or their study program. We test whether the differences are statistically significant using a two-sided Wilcoxon signed-rank test for paired samples (with the exact approximation) based on the respective group-level averages and pool all treatments to ensure sufficient statistical power.[26] We do not find any statistically significant differences ($p \geq 0.254$). Result 2 suggests that the composition effect severely hinders the successful implementation of non-grand, yet still profitable institutions in the case of heterogeneity in the individual willingness to implement such institutions. We find that a significant share of subjects, which is comparable across group sizes, always votes against the implementation of institutions in the presence of single or very few nonparticipants. This raises the question of whether this behavior can be rationalized with social preferences. In the following, we compute the Fehr-Schmidt-parameter α_i which would be in line with this behavior and compare the results to prior research.

Table A.3: Differences between Nay-Sayers and Remaining Subjects

	Nay-Sayers	Rest
Age (avg.)	24.56	24.96
Male (in %)	38.78	47.50
Study program: Economics or Business (in %)	34.00	37.25
Control questions: number of incorrect answers (avg.)	8.10	7.52

Notes: Comparison of age, sex, and study program as well as performance in control questions (number of total incorrect attempts over all 12 control questions) between nay-sayers (subjects who always vote against the implementation of an almost-grand institution when they participate in one) and the remaining subjects who vote for the implementation of such institutions at least once. Observations are pooled across all treatments.

To do so, we consider institutions of size of 75% and compute α_i such that the utility of voting in favor of the institution is smaller than the utility of voting against the implementation. This gives a lower bound for α_i (rejecting a relatively larger institution requires an

[26]If we conduct the tests separately by treatment we find some significant effects for certain treatments (nay-sayers are less likely to be male ($p = 0.078$) or study Economics or Business ($p = 0.043$) for *Treat-L* and have less incorrect answers ($p = 0.031$) for *Treat-M*), but no consistent picture emerges and none of them is significant at a conventional level if we apply the Bonferroni correction for the 12 tests (3 per dimension). Analyses are available upon request.

even higher α_i). For this, we use equation (9) from Appendix E and assume that in case of no implementation, subjects believe everybody not to contribute anything, while nonparticipants are believed not to contribute anything in case of a successful implementation. In *Treat-S* it is optimal for participants to reject an almost-grand institution if $\alpha_i \geq 1.91$, while this threshold slightly increases to $\alpha_i \geq 2.23$ and $\alpha_i \geq 2.33$ in *Treat-M* and *Treat-L* respectively.[27] These parameters do not appear unrealistically strong at a first glance. To shed more light on this, we compare the share of subjects whose behavior can be rationalized with the parameters above in our experiment with the results in Blanco, Engelmann and Normann (2011). Blanco, Engelmann and Normann (2011) find that roughly 19 out of 72 subjects exhibit behavior which implies $\alpha_i \geq 1.5$ in their experiment. This result (26.39%) is very comparable to the share of nay-sayers in our experiment, i.e., 18.60% in *Treat-S*, 24.49% in *Treat-M*, and 15.93% in *Treat-L*.

[27] If we use the actual average contributions for the respective cases as proxies for beliefs, the thresholds are $\alpha_i \geq 1.24$ for *Treat-S*, $\alpha_i \geq 1.94$ for *Treat-M*, and $\alpha_i \geq 1.70$ for *Treat-L*.

D Behavior in Every Group

Figures A.12, A.13, and A.14 give an overview of the key outcomes for every group in every round of our experiment. Player-level overviews are available upon request.

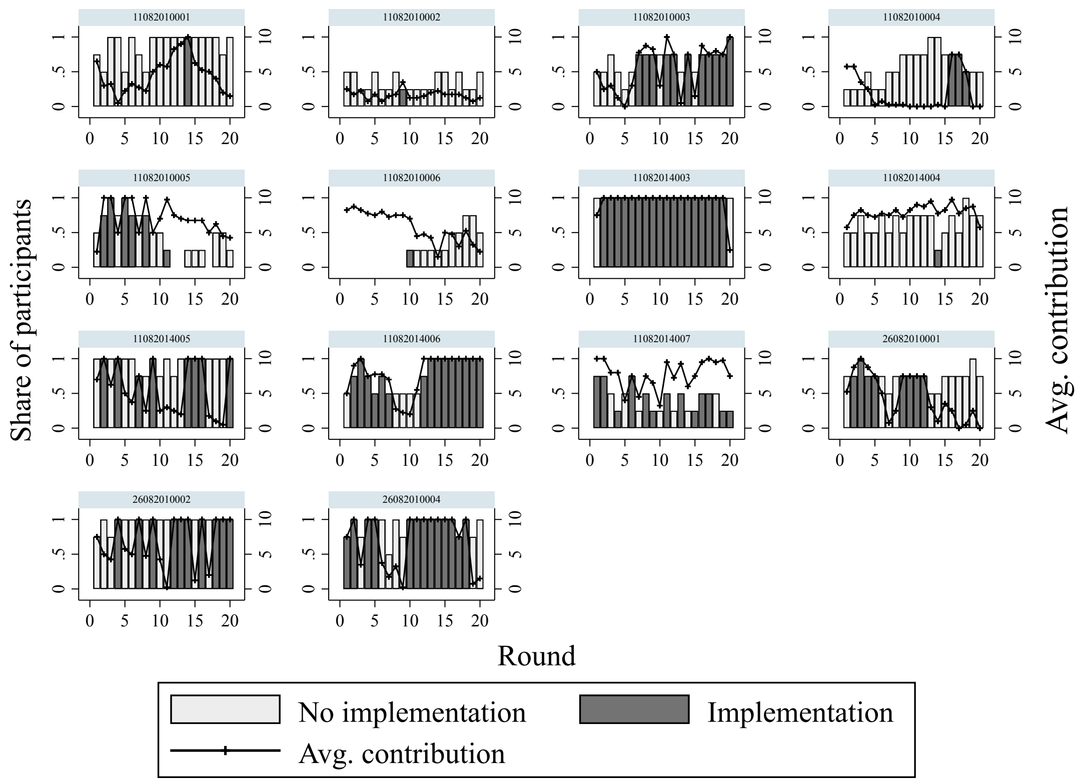


Figure A.12: Behavior of Groups in *Treat-S*

Notes: Behavior of each group in *Treat-S* for all rounds, including the share of participants, the final implementation decision, and average contributions per player. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players.

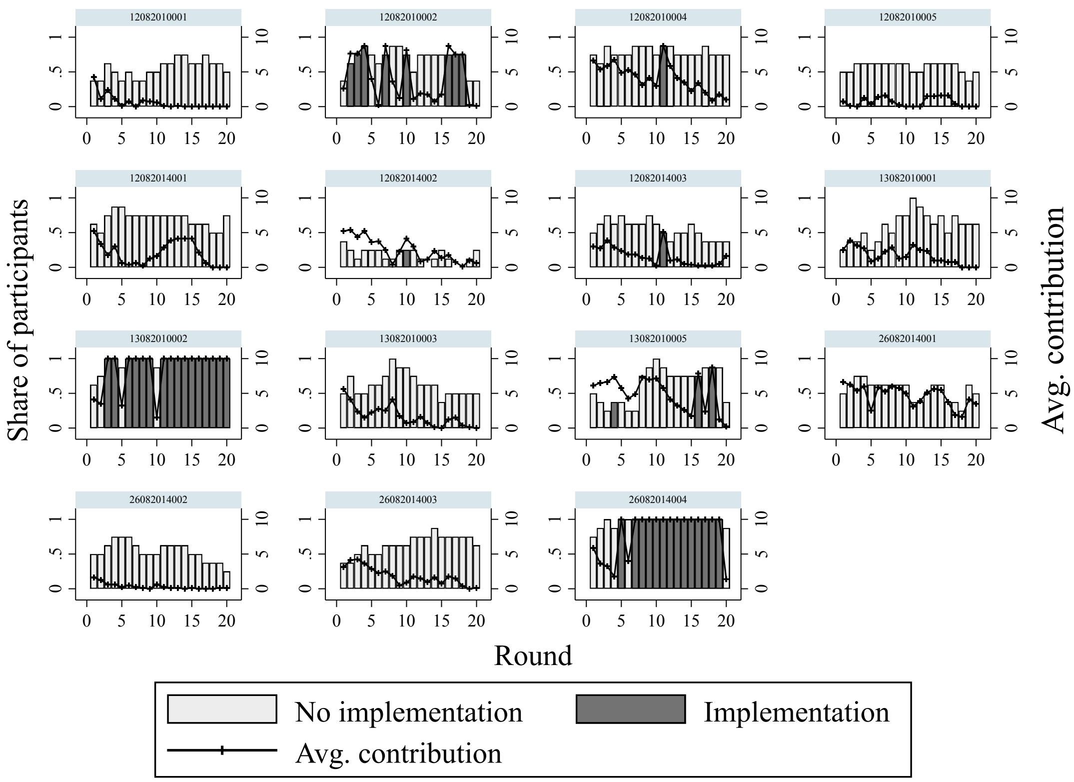


Figure A.13: Behavior of Groups in *Treat-M*

Notes: Behavior of each group in *Treat-S* for all rounds, including the share of participants, the final implementation decision, and average contributions per player. Group size is 4 (*Treat-M*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level.

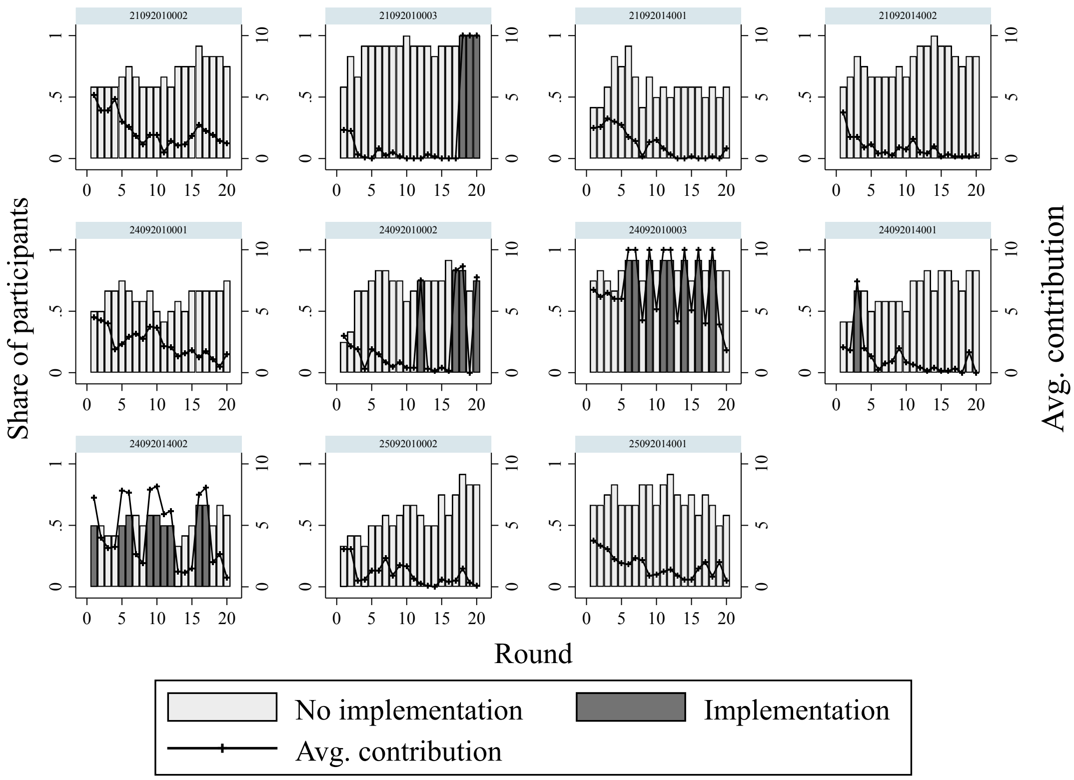


Figure A.14: Behavior of Groups in *Treat-L*

Notes: Behavior of each group in *Treat-L* for all rounds, including the share of participants, the final implementation decision, and average contributions per player. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data is pooled on the treatment level.

E Equilibrium Analysis

In this section we describe the pure-strategy equilibria in the institution formation game and how they depend on group size. The game as described in Section 2 is a n-player three-stage game in which players have perfect information about the current stage and all previous stages. We therefore consider the set of subgame-perfect Nash equilibria. A pure strategy in this game is given by (i) a participation decision $P_i \in \{0, 1\}$, (ii) an implementation decision I_i conditional on the number of participants $p = |P|$ with $P = \{i, P_i = 1\}$, and (iii) a contribution decision conditional on the number of members $s = |S|$.

We consider two separate classes of preferences: when players only care about their own monetary payoff ("money-maximizing preferences") and when players also care about others ("social preferences"). For the latter type of preferences we follow Kosfeld, Okada and Riedl (2009) and assume that players dislike inequality in payoffs based on the model by Fehr and Schmidt (1999). While there exist other social preference models emphasizing, e.g., the role of negative and positive reciprocity, our analysis based on payoff inequality does address such considerations at least qualitatively to the extent that members suffering from disadvantageous payoff inequality (the key driver in the analysis) consider the non-participation of other players as intentionally unkind. The following analysis is closely related to Kosfeld, Okada and Riedl (2009).

E.1 Money-Maximizing Preferences

Using backwards induction, all members contribute their endowment w, since members' commitments are strictly enforced. Nonmembers, however, are free in their contribution decision and will choose to contribute nothing due to $\delta/n < 1$. Members of an institution of size s therefore earn $\frac{\delta s w}{n} - c$. If no institution is implemented, nobody contributes and everybody earns w. Participants thus have an incentive to implement the institution in the implementation stage if and only if

$$\frac{\delta}{n} s w - c > w. \tag{5}$$

Let s^* denote the smallest integer s satisfying this trade-off, characterizing the minimal institution which is profitable to implement. Due to our assumption that $\delta w - c > w$, $s^* \leq n$. Because $\delta/n < 1$, $s^* \geq 2$. As the left-hand side of condition (5) is strictly increasing in s, s^* exists and is unique. Thus, players can, in principle, overcome the social dilemma problem. For clarity in exposition, we use the term *institutional equilibrium* in the following to refer to a subgame perfect Nash equilibrium in which an institution is implemented on the equilibrium path.

Proposition 1. *If players have money-maximizing preferences, the institution formation game has an institutional equilibrium with s players becoming members if and only if $s \geq s^*$. There also exists an equilibrium in which no institution is formed and nobody contributes to the public good regardless of the number of participants.*

The proof is identical to the proof of Proposition 1 in Kosfeld, Okada and Riedl (2009), taking into account the modified payoff structure for the definition of s^* as derived in equation (5). From this, it follows that the grand institution with $s = n$ always is an institutional equilibrium ($s^* \leq n$ exists by assumption).

Proposition 1 shows that there exist multiple institutional equilibria if $s^* < n$. Players need to coordinate on which institution to form. Strictness as an equilibrium refinement solves this equilibrium selection problem. A subgame perfect Nash equilibrium is called strict if it is strict on the equilibrium path, i.e., players have a unique best response in each stage of the game. This refinement criterion coincides with the concept of a stable coalition in D'Aspremont et al. (1983). If we apply strictness as an equilibrium refinement, only the institutional equilibrium with $s = s^*$ members survives. To see this, consider any $s > s^*$. In the implementation stage, any single participant is strictly better off by not participating (and not contributing in stage 3) and letting the remaining $s - 1 \geq s^*$ participating players implement the institution. Thus, there exists a profitable deviation. This is not possible if and only if $s = s^*$, because in this case non-participation by a single player leads to the rejection of the institution in the second stage. The threshold s^* thus characterizes the unique strict institutional equilibrium.

Rearranging condition (5) (and assuming equality), we see that the relative size s^*/n of the unique strict institutional equilibrium is independent of group size n, up to the possible rounding to the next integer:[28]

$$\frac{s^*}{n} = \frac{w + c}{\delta w} \tag{6}$$

This follows directly from the payoff structure. As group size n increases, the public good becomes individually less attractive due to δ/n. This lowers members' payoffs in

[28]In the experiment we choose parameters such that $s^*/n = 0.5$ for all n.

an institution, which is offset by a proportionate increase in the number of members. Proposition (2) summarizes this.

Proposition 2. *If players have money-maximizing preferences, the institution formation game has a unique strict institutional equilibrium with exactly s^* players becoming members of the institution. The relative size of this institution is independent of group size, up to rounding to the next integer.*

Proof. Consider institution S and the payoff of players in the participation stage:

$$u_i = \begin{cases} \frac{\delta}{n} sw - c & \text{if } i \in S \\ w + \frac{\delta}{n} sw & \text{if } i \notin S \end{cases} \tag{7}$$

if $s \geq s^*$, and $u_i = w$ for $s < s^*$. Let $U_i^p(s)$ ($U_i^{np}(s)$) be the utility of participant (nonparticipant; only defined for $s < n$) i of an institution of size s, with $U_i = u_i$.

Suppose that $s > s^*$ and that an institution is implemented. Further suppose that participant i deviates. Since the remaining $s-1$ participants still implement the institution with $s-1$ members, $U_i^{np}(s-1) - U_i^p(s) = w\left(1 - \frac{\delta}{n}\right) + c > 0$. Participant i has an incentive to deviate and $s > s^*$ cannot be supported by a Nash equilibrium in the participation stage.

Suppose that $s = s^*$. If participant i deviates, the institution is not implemented. By the definition of s^*, $U_i^{np}(s-1) - U_i^p(s) < 0$. Moreover, no nonparticipant has an incentive to join the institution, because this reduces the payoff from $w + \frac{\delta}{n} s^* w$ to $\frac{\delta}{n}(s^* + 1)w - c$. Consequently, every action profile with exactly s^* participants in the participation stage is a strict Nash equilibrium.

Suppose that $s = s^* - 1$. Any single nonparticipant has an incentive to participate as this leads to the implementation of the institution and increases her payoff from w to $\frac{\delta}{n} s^* w - c$. Last but not least, consider $s < s^* - 1$. Any action profile is a Nash equilibrium, but not a strict one. □

E.2 Social Preferences

We now consider social preferences. Notice at first that if $s^* < n$ there exists payoff inequality between members and nonmembers of the institution. Participants anticipating this payoff inequality might object to it and vote against the implementation of an institution of size $s < n$. The grand institution with $s = n$, however, does not generate any payoff inequality. Sufficiently strong social preferences thus favor the grand institution as the unique institutional equilibrium.

In the following, we assume that players have social preferences based on the model of Fehr and Schmidt (1999) with player i's utility U_i defined as follows:

$$U_i = u_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} max\left\{(u_j - u_i), 0\right\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} max\left\{(u_i - u_j), 0\right\}. \qquad (8)$$

The parameter α_i (β_i) captures the individual-specific aversion towards payoff inequality when player i is behind (ahead). In line with Fehr and Schmidt (1999), we assume that $\beta_i \leq \alpha_i$ and $0 \leq \beta_i \leq 1$ for all i.

If no institution is implemented and nobody contributes to the public good, there is no payoff inequality and every player earns w. Suppose for the moment that players have limited compassion and do not contribute to the public good when they are nonmembers, i.e., $\beta_i < 1 - \frac{\delta}{n}$ for all i. Then, the following equation has to be fulfilled for every participant of an institution of size s such that the institution is successfully implemented:

$$\frac{\delta}{n} sw - c - \frac{\alpha_i}{n-1}(n-s)(w+c) > w. \qquad (9)$$

Consider the grand institution $s = n$. Suppose that at least two players have sufficiently high α_i such that condition (9) is violated for $s = n-1$. Regardless of which player deviated from the grand institution in the first place, at least one player will therefore object to the implementation of an institution with $s = n - 1$. That is, the grand institution is a strict equilibrium.[29] If this condition is even violated for $n - 1$ players, no group of players is willing to implement any institution smaller than n, as the left-hand side of (9) increases with s. The grand institution then is the unique institutional equilibrium.

Suppose next that some players are sufficiently compassionate to contribute as nonmembers, i.e., $\beta_i > 1 - \frac{\delta}{n}$ for some i. In this case payoff inequality between members and nonmembers decreases and participants need to dislike disadvantageous inequality even more such that the grand institution becomes the unique strict equilibrium. Proposition 3 summarizes this.

Proposition 3. *Suppose that players have social preferences as specified in equation (8) and let* $C = \left\{i | \beta_i > 1 - \frac{\delta}{n}\right\}$. *The grand institution is a strict institutional equilibrium if*

[29] It takes two players, since a single player with sufficiently high α_i might be the one deviating from the grand institution.

and only if there exist at least two players with $\alpha_i > \tilde{\alpha}$, where

$$\tilde{\alpha} = \begin{cases} \frac{(n-1)\left(w\left(\frac{\delta}{n}(n-1)-1\right)-c\right)}{w+c} & \text{if } C = \emptyset, \\ \frac{(n-1)(w(\delta-1)-c)}{c} & \text{if } C \neq \emptyset. \end{cases} \tag{10}$$

If at least $n-1$ players satisfy $\alpha_i > \tilde{\alpha}$, implementation of the grand institution is the unique institutional equilibrium.

The proof of Proposition 3 builds on the following Lemma.[30]

1. An institution S is called *profitable* if the payoff of all participants $i \in S$ is strictly greater than if no institution was implemented and everybody contributed nothing (i.e., earning w).

2. A participant i of institution S is called *pivotal* if S is profitable but $S - \{i\}$ is not.

Lemma 1. *The grand institution with $s = n$ is implemented in a strict subgame perfect equilibrium if and only if every member is pivotal.*

Proof. First, suppose that every participant of the grand institution with $s = n$ is pivotal. Then by the definition of *pivotal* the grand institution is implemented in a strict subgame perfect equilibrium.

Second, suppose that the grand institution is implemented in a strict subgame perfect equilibrium. We show that every participant $i \in S$ is pivotal for this institution. For this, suppose, on the contrary, that participant $i \in S$ is not pivotal, so institution $N - \{i\}$ is still profitable $\forall j \in N - \{i\}$ and is implemented. We will arrive at a contradiction by showing that the non-pivotal player benefits from not participating, i.e. $S = N$ is not a strict equilibrium. For this, let $U_i^p(s)$ ($U_i^{np}(s)$) be the utility of participant (nonparticipant; only defined for $s < n$) i of an institution of size s based on equation (8). Let $C = \left\{i|\beta_i > 1 - \frac{\delta}{n}\right\} \neq \emptyset$. We distinguish between two possible cases:

1. $i \notin C$: If player i participates in $S = N$, she receives

$$U_i^p(n) = \frac{\delta}{n} nw - c$$

[30] If $C = \left\{i|\beta_i > 1 - \frac{\delta}{n}\right\} = \emptyset$, this Lemma also applies more generally for any institution S.

If player i does not participate in $S = N$ and $i \notin C$, she will contribute 0, leading to

$$U_i^{np}(n-1) = w + \frac{\delta}{n}(n-1)w - \frac{\beta_i}{n-1}(n-1)(w+c)$$

Since $i \notin C$, $1 - \frac{\delta}{n} \geq \beta_i$ and thus

$$U_i^{np}(n-1) - U_i^p(n) = w\left(1 - \frac{\delta}{n} - \beta_i\right) + c\,(1-\beta_i) > 0$$

which is a contradiction. If $C = \emptyset$, only this case is relevant.

2. $i \in C$: If player i participates in $S = N$, she receives

$$U_i^p(n) = \frac{\delta}{n}nw - c$$

If player i does not participate in $S = N$ and $i \in C$, she will choose $g_i = w$. Consequently,

$$U_i^{np}(n-1) = \frac{\delta}{n}nw - \frac{\beta_i}{n-1}(n-1)c$$

and

$$U_i^{np}(n-1) - U_i^p(n) = c(1-\beta_i) > 0$$

which is a contradiction.

□

Proof of Proposition 3.

The grand institution does not involve inequality in payoffs and therefore is profitable by assumption. We now consider the case when player j deviates and show that every participant of the grand institution is pivotal if and only if there exist at least two players with $\alpha_i > \tilde{\alpha}$. The first part of Proposition 3 then follows from Lemma 1.

Suppose player j deviates. Consider the two cases:

1. Deviating player $j \notin C$: player j will contribute nothing to the public good. The remaining $n-1$ participants thus each earn

$$U_i^p(n-1) = \frac{\delta}{n}(n-1)w - c - \frac{\alpha_i}{n-1}(w+c) \tag{11}$$

$S = N - \{j\}$ is not profitable, i.e., $U_i^p(n-1) < w$, if and only if

$$\alpha_i > \frac{(n-1)\left(w\left(\frac{\delta}{n}(n-1)-1\right)-c\right)}{w+c} \equiv \tilde{\alpha}_1 \quad (12)$$

for at least two players. Thereby, every participant of the grand institution is pivotal if and only if $\alpha_i > \tilde{\alpha}_1$ for at least two players (one might be the one who deviates).

If $C = \emptyset$, only this case is relevant.

2. Deviating player $j \in C$: Player j will still contribute w (see proof of Lemma 1). Thus, the restriction on α_i comes from:

$$U_i^p(n-1) = \frac{\delta}{n}nw - c - \frac{\alpha_i}{n-1}c < w \quad (13)$$

$$\alpha_i > \frac{(n-1)\left(w(\delta-1)-c\right)}{c} \equiv \tilde{\alpha}_2 \quad (14)$$

Clearly $\tilde{\alpha}_2 > \tilde{\alpha}_1$. Thus, having at least two players with $\alpha_i > \tilde{\alpha}_2$ is the stricter condition when $C \neq \emptyset$.

Now consider the second part, i.e., uniqueness. This requires that $\forall s < n$ at least one participant votes against the implementation, i.e., the utility as a member of the institution of size s is smaller than the utility without any institution (w by assumption).

Suppose at first that $C = \emptyset$. In this case, nonmembers do not contribute to the public good and a members' payoff strictly increases in s. Moreover, inequality also strictly decreases in s. It follows that the utility of a member of any institution with $s \leq n-1$ is maximized for $s = n-1$. If $\alpha_i > \tilde{\alpha}_1$ for $n-1$ players, clearly any institution with $s \leq n-1$ will always be rejected by at least a single participant. The grand institution thus is the unique institutional equilibrium.

Now suppose $C \neq \emptyset$. While a member's monetary payoff does not necessarily increase in s anymore[31], it still holds that the utility of a member of any institution with $s \leq n-1$ is maximized for $s = n-1$ and nonparticipating player $j \in C$. Clearly, the monetary payoff of any institution with $s \leq n-1$ is (weakly) smaller than full contribution by everyone. Moreover, $\sum_{j \neq i} |u_i - u_j| \geq c$, when there is just a single nonmember who still contributes

[31] To be precise, it can actually decrease in s. For example, suppose that $n = 4$ and $C = \{1, 2, 3\}$. Clearly, the institution $S = \{1, 4\}$ leads to a higher payoff for members than the institution $S = \{1, 2, 3\}$, since $i = 2, 3$ contribute regardless of their institution membership while $i = 4$ does not.

w. It follows that if $n-1$ players are sufficiently averse to inequality to reject this "best-case" scenario implied by Equation (13), i.e., $\alpha_i > \tilde{\alpha}_2$, any institution with $s < n$ will also be rejected. $\square$

Finally, consider the effect of group size if players have social preferences. In the main text, we distinguish between three possible channels: preference effect, threshold effect, and composition effect. First, the preference effect refers to the conjecture that social preferences themselves might play a weaker role when group size increases (Schumacher et al., 2017; Alós-Ferrer, García-Segarra and Ritschel, 2022). Clearly, this could affect whether there are at least two players with $\alpha_i > \tilde{\alpha}$. Second, while the payoffs of members and non-members (and thus payoff inequality between them) in an institution of the same relative size do not depend on group size (cf. Proposition 2), larger group sizes allow for larger non-grand institutions which features lower (average) inequality and higher profitability. This increases the threshold $\tilde{\alpha}$ and we refer to it as the threshold effect. Formally, $\tilde{\alpha}$ increases in n; depending on the distribution of α_i, the increase in $\tilde{\alpha}$ can affect the uniqueness and strictness of the grand institution as an equilibrium. Third, for a given distribution of α_i in a population, the likelihood that at least subjects have $\alpha_i > \tilde{\alpha}$ increases with group size. We refer to this as the composition effect.

In sum, the theoretical analysis makes two important observations. First, if players have money-maximizing preferences, the relative size of the unique strict institutional equilibrium, and thus efficiency, does not vary with group size. This is a very clear prediction on what should happen in the absence of social preferences. Second, social preferences, based on members disliking payoff inequality between themselves and nonmembers, favor the grand institution. The effect of group size on institution formation then depends on the interplay of the preference, threshold, and composition effect.

Recent Issues

No. 401	Kevin Bauer, Oliver Hinz, Michael Kosfeld, Lena Liebich	Decoding GPT's Hidden 'Rationality' of Cooperation
No. 400	Andreas Hackethal, Philip Schnorpfeil, Michael Weber	Households' Response to the Wealth Effects of Inflation
No. 399	Raimond Maurer, Sehrish Usman	Dynamics of Life Course Family Transitions in Germany: Exploring Patterns, Process and Relationships
No. 398	Pantelis Karapanagiotis, Marius Liebald	Entity Matching with Similarity Encoding: A Supervised Learning Recommendation Framework for Linking (Big) Data
No. 397	Matteo Bagnara, Milad Goodarzi	Clustering-Based Sector Investing
No. 396	Nils Grevenbrock, Alexander Ludwig, Nawid Siassi	Homeownership Rates, Housing Policies, and Co-Residence Decisions
No. 395	Ruggero Jappelli, Loriana Pelizzon, Marti Subrahmanyam	Quantitative Easing, the Repo Market, and the Term Structure of Interest Rates
No. 394	Kevin Bauer, Oliver Hinz, Moritz von Zahn	Please Take Over: XAI, Delegation of Authority, and Domain Knowledge
No. 393	Michael Kosfeld, Zahra Sharafi	The Preference Survey Module: Evidence on Social Preferences from Tehran
No. 392	Christian Mücke	Bank Dividend Restrictions and Banks' Institutional Investors
No. 391	Carmelo Latino, Loriana Pelizzon, Max Riedel	How to Green the European Auto ABS Market? A Literature Survey
No. 390	Kamelia Kosekova, Angela Maddaloni, Melina Papoutsi, Fabiano Schivardi	Firm-Bank Relationships: A Cross-Country Comparison
No. 389	Stefan Goldbach, Philipp Harms, Axel Jochem, Volker Nitsch, Alfons J. Weichenrieder	Retained Earnings and Foreign Portfolio Ownership: Implications for the Current Account Debate
No. 388	Gill Segal, Ivan Shaliastovich	Uncertainty, Risk, and Capital Growth