Journal of Economic Behavior and Organization 240 (2025) 107298

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Research Paper

The detrimental effect of group size on institution formation

Julian Detemple [a,*], Michael Kosfeld [b]

[a] Department of Business and Economics, Johannes-Gutenberg-University Mainz, Jakob-Welder-Weg 4, 55128, Mainz, Germany
[b] Faculty of Economics and Business, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323, Frankfurt am Main, Germany

ARTICLE INFO

JEL classification:
C92
D02
H41

Keywords:
Institution formation
Group size
Social dilemma

ABSTRACT

A solution for the provision of public goods is the formation of institutions that change the rules of the game, e.g., through sanctions or enforced cooperation commitments. While prior laboratory experiments document a positive effect of the opportunity to form such institutions on cooperation in small groups, groups in the field are typically much larger. We test the causal effect of group size on institution formation and show that institutions almost never form successfully as group size increases from four to just eight or twelve individuals. Prior results on the formation of institutions, such as those of Kosfeld et al. (2009), thus do not generalize to larger groups. Our findings document that individuals are less willing to be bound by the rules of the institution, while willing individuals struggle to reach a unanimous decision as group size increases.

1. Introduction

The formation of institutions that change the rules of the game (North, 1990), e.g., through sanctions or enforced commitments, is a key solution to collective action problems (Baland and Platteau, 1996; Ostrom, 1999). Examples include governance systems for communal resources, such as grazing lands in Switzerland (Ostrom, 1990) and forests in India (Agrawal and Yadama, 1997). However, often not all individuals are willing to be bound by the rules of such institutions. The formation of institutions thus depends on either rule enforcement by a higher authority or unanimous agreement among a coalition of individuals willing to change the rules only for themselves (Ostrom, 2005), despite others also benefiting from collective action. This paper focuses on the latter case. Previous laboratory research has demonstrated that the opportunity to form sanctioning or commitment institutions enhances cooperation through the formation of institutions that govern (almost) everyone (Dannenberg and Gallier, 2020). Importantly, individuals exhibit heterogeneous preferences over institutions that do not govern everyone, e.g., due to the inequality that these institutions entail when people outside the coalition free-ride on the coalition's cooperation (Kosfeld et al., 2009). If the coalition of individuals willing to change the rules for themselves encompasses (almost) everyone, institutions form more frequently than if too few individuals are willing to be bound by the rules (Kosfeld et al., 2009). Yet, these findings are based on small groups, typically around four individuals, raising the question of how group size affects institution formation. This question is of particular importance for two reasons. First, the need for institutions to solve the collection action problem may increase when groups become larger. In many collective action problems, such as the governance of communal resources, the individual benefit from cooperation decreases with group size, which has been shown to negatively affect cooperation in the absence of institutions (e.g., Isaac and Walker, 1988). Second, the effect of group size on institution formation is ambiguous when not everyone is willing to be governed by the institution. On the one hand, a larger group size naturally increases the likelihood of disagreement among individuals with heterogeneous preferences. When unanimous agreement is required, institutions that do not govern everyone would then form even less frequently in larger groups.

* Corresponding author.
E-mail addresses: julian.detemple@uni-mainz.de (J. Detemple), kosfeld@econ.uni-frankfurt.de (M. Kosfeld).

https://doi.org/10.1016/j.jebo.2025.107298
Received 21 May 2025; Received in revised form 15 October 2025; Accepted 17 October 2025
Available online 11 November 2025

On the other hand, individual preferences over forming an institution that does not govern everyone might be sensitive to group size (e.g., because individuals in larger groups are less motivated by social preferences, see Alós-Ferrer et al., 2022). This would reduce the likelihood of disagreement in larger groups and favor the formation of institutions. The effect of group size on institution formation thus depends on the combination of these channels and calls for an empirical investigation.

This paper addresses the question of how group size affects institution formation in the context of a commitment institution, i.e., an institution that changes the rules of the game through enforced cooperation commitments. The institution is implemented through unanimous agreement among the coalition of individuals willing to commit themselves. In this setting, we examine the causal effect of increasing group size from four to eight and twelve individuals in a laboratory experiment. We study how group size affects the formation of institutions and distinguish between institutions that are implemented through grand coalitions (i.e., institutions governing everyone), almost-grand coalitions, and smaller non-grand coalitions. In contrast to field settings where group size is endogenous and likely correlated with other factors that influence institution formation (e.g., the profitability of an institution), the controlled laboratory environment allows to isolate the effect of group size, while disentangling the different effects at work through the experimental design.

We find that in contrast to groups with four individuals almost no institution is formed in groups with as few as eight players. This is driven by a combination of different effects. On the one hand, the implementation of institutions when everyone is willing to commit themselves is not impacted by group size, but these grand coalitions almost never occur in larger groups. On the other hand, almost-grand coalitions are almost never implemented when group size increases beyond four players. Importantly, group size does not affect the distribution of individual preferences on the implementation of such coalitions: a great majority of subjects is in favor of implementing institutions through almost-grand coalitions, while a small yet significant share has strong preferences against them. Since group size does, however, not lower the share of individuals who object to almost-grand coalitions, group size increases the likelihood of disagreement within these coalitions. As a result, almost-grand coalitions almost never implement an institution in larger groups. Interestingly, the more frequent failure of almost-grand coalitions to form an institution is even less effective at changing the minds of the few subjects that are not willing to commit themselves in larger groups.[1] This explains why grand coalitions almost never occur when group size increases beyond four players. There are no group-size effects on the formation of smaller non-grand coalitions, which, in line with the previous literature, already fail to form in groups with just four individuals. Finally, cooperation is low in larger groups. We present evidence that cooperation is lower than what the lower institution formation rate would predict, suggesting that the negative effect of group size on cooperation in collective action problems (e.g., Isaac and Walker, 1988) also applies in the domain of institution formation.

Our paper relates to three strands in the literature. First, our findings contribute to institution formation in laboratory experiments (e.g., Ertan et al., 2009; Sutter et al., 2010; Kamei et al., 2015; Schmidt and Ockenfels, 2021). As summarized by Dannenberg and Gallier (2020), the opportunity to form institutions increases cooperation through the successful formation of institutions implemented through grand or almost-grand coalitions (e.g., Kosfeld et al., 2009; McEvoy et al., 2011). While groups with more than four individuals have been studied (e.g., Gürerk et al., 2006; Kamei et al., 2023), we are the first to analyze the effect of group size on institution formation through exogenous variation, documenting a negative effect, with almost no institution forming successfully in larger groups.

Second, we contribute to how group size influences collective action in the field (e.g., Olson, 1965; Agrawal and Goyal, 2001; Poteete and Ostrom, 2004). In this literature, group size has a non-linear effect, with medium-sized groups which are still larger than the group size we study (e.g., between 60 and 80 households in Agrawal and Goyal, 2001) exhibiting the highest success rate of organizing collective action. However, it is difficult to isolate the effect of group size from (observational) field data and "group size may be as much an indicator of institutional success as a precondition for such success" (Poteete and Ostrom, 2004, p.435). While we study collective action in comparatively smaller groups in a laboratory experiment and not the field, the controlled environment allows us to identify the causal effect of group size and disentangle the different effects at work. We show that when one compares institutions of the same profitability across different group sizes, an institution is formed much less frequently in larger groups. Our results thus highlight the question of how institutions in medium-sized groups in the field manage to overcome that negative group-size effect (i.e., how institutions in larger groups are different from institutions in smaller groups).

Third, our findings are related to the literature on group-size effects on cooperation in public goods games (e.g., Isaac and Walker, 1988; Isaac et al., 1994; Nosenzo et al., 2015; Weimann et al., 2019). In this literature, group size has no or even a positive effect on cooperation when the individual benefit from cooperation remains constant, but cooperation decreases strongly when the individual benefit from cooperation decreases when group size increases. Our setting is most similar to the latter case. In line with this literature, we find that cooperation levels are lower in larger groups than what the lower formation rate of institutions would predict. This suggests that a direct negative effect of group size on cooperation also applies in the context of institution formation.

In the next section, we describe the institution formation process in more detail and lay out a theoretical framework to provide intuition for the ambiguous effect of group size on institution formation. In Section 3, we describe the experiment. Section 4 presents our findings and Section 5 concludes.

[1] Note that players have a material incentive to "hold out" from committing to the institution hoping that other players will commit and they can free-ride.

2. Institution formation game

We adapt the n-person institution formation game analyzed in Okada (1993) and Kosfeld et al. (2009). The number of players n is common knowledge. In the spirit of North (1990), we refer to the rules of the game as an institution. Specifically, we consider an institution that gives players the opportunity to commit themselves to contributing their endowment w to a linear public good. If the institution forms successfully, the commitments are enforced automatically and the institution is implemented through the coalition of committed players. Players who do not commit themselves benefit from players' commitments and players who commit themselves benefit from (voluntary) contributions by non-committed players. This setting has been shown to be equivalent to a sanctioning institution which enforces commitments via sanctions (Kosfeld et al., 2009). Institutions are formed in three stages.

Participation stage: players simultaneously and independently indicate their willingness to commit themselves. While we refer to players who (do not) indicate their willingness as *participants* (*nonparticipants*) of the coalition of players willing to commit themselves, it is important to keep in mind that this coalition is not (yet) implemented. Consequently, participation in the coalition simply indicates the willingness to commit oneself and players are (non)participants of a *provisional* coalition, i.e., provisional participants.[2] For simplicity and to stay consistent with the terminology of Okada (1993) and Kosfeld et al. (2009), we, however, omit the "provisional"-qualifier in the following. Let s be the number of participants in the coalition. If $s \geq 1$, we say that an institution is initiated, with a coalition of size s willing to commit themselves.

Implementation stage: all players are informed about s. If an institution was initiated, all participants simultaneously and independently vote on the implementation of the institution. Only if all participants unanimously vote in favor of implementation, the institution is successfully formed and commitments become binding for the coalition of size s. Participants become *members*, nonparticipants *nonmembers*.

Contribution stage: players are informed about the implementation outcome. If implementation was successful, nonmembers decide on their respective contribution, while members contribute their endowment. The implementation of the institution comes at a fixed cost to each member.[3] If implementation was not successful, all players decide on their contribution. The marginal per capita return (MPCR) from contributing to the public good is δ/n with $0 < \delta/n < 1 < \delta$. 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 formed, assuming that players have money-maximizing preferences. By letting the MPCR depend on n, we deviate from Kosfeld et al. (2009). Importantly, as we show below, this allows us to vary group size, while holding the profitability of a coalition of the same relative size constant. Note that the institution formation game allows players to reduce the strategic uncertainty about other players' willingness to cooperate (Steimanis et al., 2025).

Let us introduce some further notation. A player's payoff u_i depends on whether an institution is formed. Let S denote the set of $s = |S|$ members of the coalition that implemented the institution, and let $c > 0$ be the institution formation cost for each member.

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

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

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

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

Eqs. (1) and (2) highlight that nonmembers of a coalition benefit from members' contributions and do not have to bear the institution formation cost c. In addition, they are free in their contribution decisions. This allows them to free-ride on members' commitments, leading to payoff inequality between members and nonmembers. We assume that $\delta w - c > w$, i.e., it is socially optimal for the grand coalition with $s = n$ to form the institution.

We now provide some intuition on the institution formation process and how group size can affect it through different channels. In Appendix E, we expand on this and provide a more detailed equilibrium analysis.

First, suppose that players have (homogeneous) money-maximizing preferences, i.e., maximize u_i. We call a coalition *profitable* if members' payoff is higher than without any institution. If a coalition of size s implements the institution, members contribute w and nonmembers contribute nothing due to $\delta/n < 1$. If no institution is formed, nobody contributes and everybody earns w. A coalition of size s is thus profitable for a member iff

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

Since $\delta w - c > w$, sufficiently large coalitions are profitable. First, notice that for a fixed δ (as well as w and c), the relative size of a coalition s/n alone determines whether a coalition is profitable to implement. This allows to vary group size and compare coalitions of the same relative size while holding constant the profitability and thus material incentives of forming an institution. Let s^* denote the smallest integer s satisfying this trade-off, characterizing the smallest coalition that is profitable to implement. Clearly, also the

[2] We thank an anonymous reviewer for pointing this out.

[3] From an applied perspective, this cost may capture, e.g., resources needed to enforce commitments.

relative size of the smallest profitable coalition, s^*/n, is independent of group size.[4] In Appendix E, we show that a coalition of size s^*/n is the unique strict equilibrium with an institution formed on the equilibrium path.[5] This allows to derive a clear hypothesis: if players were only motivated by monetary concerns, we should not observe any group-size effects in our experiment.

Second, consider heterogeneous preferences, with some players having preferences against the implementation of non-grand coalitions with $s < n$ due to the payoff inequality between members and nonmembers (Kosfeld et al., 2009; Dannenberg and Gallier, 2020). The inequality in payoffs between a member (M) and a nonmember (NM) is determined by Eq. (1) for $s < n$:[6]

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

Notice that the payoff inequality does not depend on n. Using the model of Fehr and Schmidt (1999), let α_i denote a player's aversion to disadvantageous payoff inequality. If players are sufficiently averse to payoff inequality, all coalitions other than the grand coalition (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 coalition to be the unique equilibrium with an institution formed on the equilibrium path: there need to be (1) sufficiently many players with (2) sufficiently strong preferences against non-grand coalitions (i.e., $\alpha_i > \tilde{\alpha}$).

With heterogeneous preferences, the effect of group size on institution formation, in particular the implementation of non-grand coalitions, becomes ambiguous. First, group size mechanically increases the likelihood that there will be players with $\alpha_i > \tilde{\alpha}$ in a group. This suggests that fewer non-grand coalitions should be implemented when group size increases. Second, and operating in the other direction, preferences themselves might be sensitive to group size (Schumacher et al., 2017; Alós-Ferrer et al., 2022), i.e., individuals might object less to the *same* inequality in larger groups.[7] This would translate into lower α_i. Third, 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 coalitions (e.g., comprising 11 out of 12 compared to 3 out of 4 players) which are more profitable and generate lower inequality. The overall effect of group size on institution formation thus depends on the combination of all three effects and calls for an empirical investigation.

Summing up, if players have money-maximizing preferences, group size does not affect the institution formation process and groups should implement coalitions of the same relative size, i.e., s^*/n. Any finding other than that implies the presence of heterogeneous preferences in small and/or large groups. Notice that this setup then allows us to disentangle the influence of the three effects spelled out above. Because coalitions of the same *relative* size feature the same profitability, we can compare implementation decisions across group sizes while holding constant profitability. If preferences against inequality in non-grand coalitions, for example, play a weaker role in larger groups, we should observe that individual players are more likely to vote in favor of profitable non-grand coalitions of the same relative size.

3. Experimental design

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 et al. (2009). Before subjects could start with the experiment, they had to answer a set of control questions correctly.

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 et al. (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 coalition size which is profitable to implement corresponds to $s^* = 2, 4, 6$ respectively for the three treatments.[8]

A total of 56, 120, and 132 participants participated in *Treat-S* (14 groups), *Treat-M* (15 groups), and *Treat-L* (11 groups), respectively. All data were collected from a student sample in 2020. Additional details on the experiment are provided in Appendix A.

4. Results

First, consider how group size affects overall institution formation.[9] Panel A in Table 1 displays the initiation and implementation rate of coalitions across treatments, while Panel B shows average contributions to the public good per player.

Group size strongly influences the institution formation process. While institutions are initiated in (almost) all rounds, i.e., in more than 95 % of the rounds a coalition of minimum size one is initiated, the implementation rate drops from 41 % in *Treat-S* to 15 % in *Treat-M* ($p = .005$) and 11 % in *Treat-L* ($p = .004$), with no significant difference between *Treat-M* and *Treat-L*.[10] Because institutions

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

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

[6] For simplicity, we assume here that nonmembers do not contribute to the public good. We relax this assumption in Appendix E. If (some) nonmembers also contribute to the public good, Eq. (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.

[7] While preferences against non-grand coalitions might also become stronger with group size, the evidence cited above points to such preferences playing a weaker role in larger groups — albeit in different strategic settings.

[8] This is also in line with the experiment in Kosfeld et al. (2009) for *Treat-S*.

[9] Appendix C shows that all results are robust to accounting for learning. Data and replication code are available at Detemple and Kosfeld (2025).

[10] Tests are based on group averages. The respective p-values refer to pairwise two-sided Mann-Whitney-U tests with the exact p-values, unless noted otherwise.

Table 1
Institution formation & contributions.

A. Institution Formation	*Treat-S*	*Treat-M*	*Treat-L*
Formation Rate (in %)	38.93	14.67	11.36
Initiation Rate (in %)	95.71	99.00	100.00
Implementation Rate (in %)	40.67	14.81	11.36
Relative Size of Implemented Coalition (in %)	82.80	90.01	77.00
B. Public Good Contributions			
Contribution per Player	5.99	3.17	2.40

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

Table 2
Contributions by implementation outcome.

A. Successful Implementation	*Treat-S*	*Treat-M*	*Treat-L*
Frequency (in %)	38.93	14.67	11.36
Contribution per Player	9.16	9.24	8.65
From Members	8.28	9.01	7.70
From Nonmembers	0.88	0.24	0.95
B. Failed Implementation			
Frequency (in %)	56.79	84.33	88.64
Contribution per Player	3.70	2.15	1.60
From Participants	2.41	1.43	1.06
From Nonparticipants	1.29	0.71	0.54

Notes: Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data are pooled on the treatment level. We omit the case when no institution was initiated, since an institution was almost always initiated across group sizes (cf. Table 1). Contributions are normalized by dividing the average total contribution from all players, members, participants, etc. by the number of total players in the respective treatment.

are almost always initiated, results for the overall formation rate (i.e., the product of the initiation and the implementation rate) are almost identical to the implementation rate. Although the relative size of the implemented coalitions increases (decreases) from 83 % of players to 90 % (77 %) in *Treat-M* (*Treat-L*), these differences are not statistically significant ($p \geq 0.805$). However, it is important to note that across all treatments the average relative size is substantially larger than what money-maximizing preferences would predict with our parametrization, i.e., $s^*/n = 0.5$ (cf. Sections 2 and 3). This is in line with Kosfeld et al. (2009) and suggests that at least some players have preferences against the inequality associated with non-grand coalitions.[11]

Turning to Panel B shows that while contributions to the public good are at 5.99 points (out of 10 points) per player in *Treat-S*, contributions in larger groups are low: 3.17 points in *Treat-M* ($p = .008$) and even just 2.40 points in *Treat-L* ($p = .001$). At this point, it is important to keep in mind that players are free in their contribution decisions if they are nonmembers or an institution failed to form. Participants might, for example, decide to contribute their endowment even when an institution fails to form. Table 2 therefore shows contributions to the public good separately in case of a successful and a failed implementation.[12] Panel A highlights that the average contribution per player in case of a successful implementation is very similar across group sizes ($p \geq 0.632$), with nonmembers mostly free-riding on members' commitments.[13] Panel B documents that contributions are low when an institution fails to form, i.e., players do not voluntarily contribute their endowment to the public good on average. Moreover, the average contribution per player in case of a failed implementation is substantially smaller in larger groups ($p = .029$ for the difference between *Treat-S* and *Treat-M*, $p = .005$ for the difference between *Treat-S* and *Treat-L*). Importantly, this holds for both participants and nonparticipants. This suggests that overall contributions are lower than what the lower institution formation rate (and slightly different relative sizes) would predict. Where does this come from? Since we do not include a treatment without the possibility to form institutions, we

[11] As an important side-result, *Treat-S* largely replicates the results of Kosfeld et al. (2009).

[12] To enable comparisons across different group sizes (and numbers of participants, members, etc.), the table normalizes contributions and divides the respective total contributions (coming from participants, members, etc.) by the total number of players in the respective treatment. The contribution per player from, for example, members thus is equivalent to the endowment (10 points) times the average relative size of implemented coalitions.

[13] The change in the average contribution per player from members across groups corresponds to the change in the relative size of implemented institutions (cf. Table 1).

Table 3
Institution formation by coalition size.

	Treat-S	*Treat-M*	*Treat-L*
A. Share of Initiated Coalitions (in %)			
Grand Coalition	36.94	11.78	2.27
Almost-Grand Coalition	26.87	31.99	42.73
Other	36.19	56.23	55.00
B. Implementation Rate (in %)			
Grand Coalition	60.61	85.71	60.00
Almost-Grand Coalition	44.44	10.53	12.77
Other	17.53	2.40	8.26

Notes: A grand coalition requires everyone to participate, i.e., $s/n = 1$. An almost-grand coalition refers to $s/n \in [0.75, 1)$, with other referring to $s/n \in [0, 0.75)$. Group size (n) is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data are pooled on the treatment level.

cannot isolate the effect of institution formation on contributions. Table 2, however, suggests that multiple effects may be at work. On the one hand, seeing institutions fail to form more frequently could affect participants' willingness to contribute (e.g., through more disappointment). We do find some suggestive evidence in favor of this. Across all treatments, we observe that contributions per participant in case of a failed implementation are lower in groups with more failed implementations (Spearman's $\rho = -0.3644$ with $p = .021$). On the other hand, larger groups are known to cooperate less when the individual benefit from cooperation decreases with group size (Isaac and Walker, 1988). For this, consider nonparticipants, who are not interested in an institution in the first place and should thus not be affected by the formation failure. As Panel B in Table 2 shows, nonparticipants contribute less in larger groups in case of a failed implementation. This speaks to the negative group-size effect documented in, e.g., Isaac and Walker (1988).

Having established that group size affects overall institution formation through a lower implementation rate, we now turn to why group size has this effect.

The overall implementation rate can be decomposed into the average of the implementation rate of initiated coalitions of each size s (Panel B in Fig. 1), weighted with the frequency with which these coalitions are initiated (Panel A in Fig. 1). Starting with analyzing the implementation rate of different coalition sizes (Panel B), group size only impacts the implementation of so-called "almost-grand" coalitions, i.e., non-grand coalitions with at least 75 % of the group participating.[14] For example, when $n-1$ players indicate their willingness to commit themselves, implementation rates drop from 0.44 ($N_G = 12$) in *Treat-S* to 0.12 ($N_G = 9$) in *Treat-M* ($p = .034$) and 0.28 ($N_G = 8$) in *Treat-L* ($p = .057$). By contrast, the implementation rate of both grand coalitions and profitable coalitions with less than 75 % of the group participating is similar or, if at all, even higher in *Treat-M* for grand coalitions. Differences are, however, not statistically significant ($p \geq 0.338$).[15] Moreover, Panel A indicates that the lower implementation rate of almost-grand coalitions is further exacerbated by fewer initiated grand coalitions in larger groups (i.e., coalitions that are actually implemented in larger groups). Conditional on initiating an institution, the share of grand coalitions decreases from 37 % in *Treat-S* to 12 % in *Treat-M* ($p = .011$) and even 2 % in *Treat-L* ($p = .002$).

Consequently, Fig. 1 suggests that the negative group-size effect on institution formation is driven by different institution formation dynamics for different coalition sizes, in particular grand coalitions and almost-grand coalitions. Table 3 aggregates the institution formation process on that level.[16] Panel B confirms the insights based on Fig. 1: group-size negatively impacts the implementation of almost-grand coalitions. The implementation rate for almost-grand coalitions decreases from 44 % ($N_G = 12$) in *Treat-S* to 11 % ($N_G = 13$) in *Treat-M* ($p = .011$) and 13 % ($N_G = 10$) in *Treat-L* ($p = .025$). Moreover, Panel A highlights again that larger groups are less likely to initiate grand coalitions and instead initiate almost-grand coalitions (27 % in *Treat-S* compared to 32 % in *Treat-M* and 43 % in *Treat-L*). Group size thus affects institution formation by both reducing the frequency of initiated grand coalitions (i.e., coalitions for which implementation is most successful and does not differ across group sizes) and the implementation rate of almost-grand coalitions (i.e., coalitions that are actually more frequently initiated in larger groups).[17] We further explore possible channels for these effects in the following.

Consider the implementation of almost-grand coalitions in the presence of heterogeneous preferences over them, e.g., due to heterogeneous social preferences. As outlined in Section 2, multiple effects may be at work. First, if preferences themselves are not

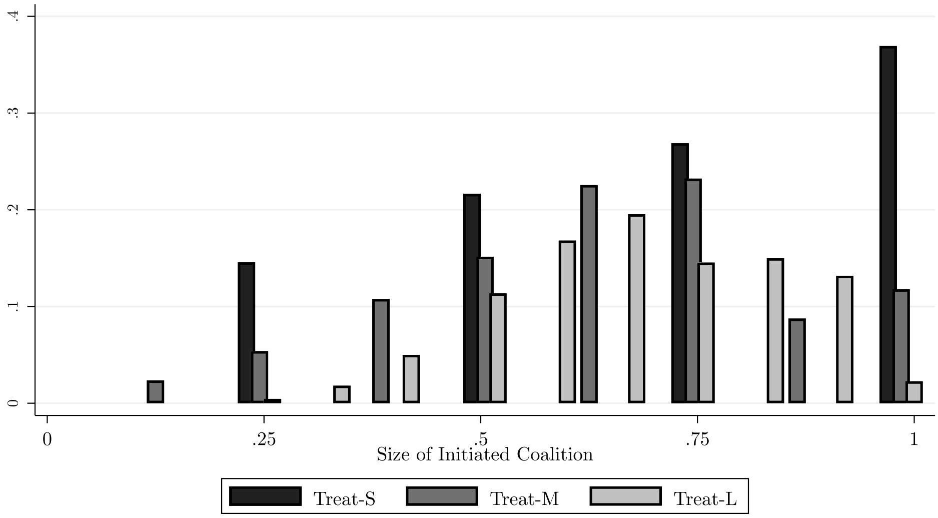
[14] It is important to remember that groups could only vote on the coalitions they initiated. Groups were thus not given the same choices, with the number of observations for a given coalition size in Panel B depending on how frequently that coalition size was initiated (Panel A). While this does not affect that we are comparing implementation outcomes for coalitions of the same relative size and thus profitability, it does affect the power of the associated statistical tests for some coalition sizes. We therefore additionally denote the number of groups (i.e., independent observations) for each test for each treatment as N_G.

[15] For the grand (50 %) coalition, $N_G = 11$ ($N_G = 11$) for *Treat-S*, $N_G = 5$ ($N_G = 11$) for *Treat-M*, and $N_G = 2$ ($N_G = 6$) for *Treat-L*. Smaller coalitions are not profitable to implement. Figs. D.1–D.3 in Appendix D document the institution formation in each single group for all treatments.

[16] We also include other non-grand coalitions for completeness.

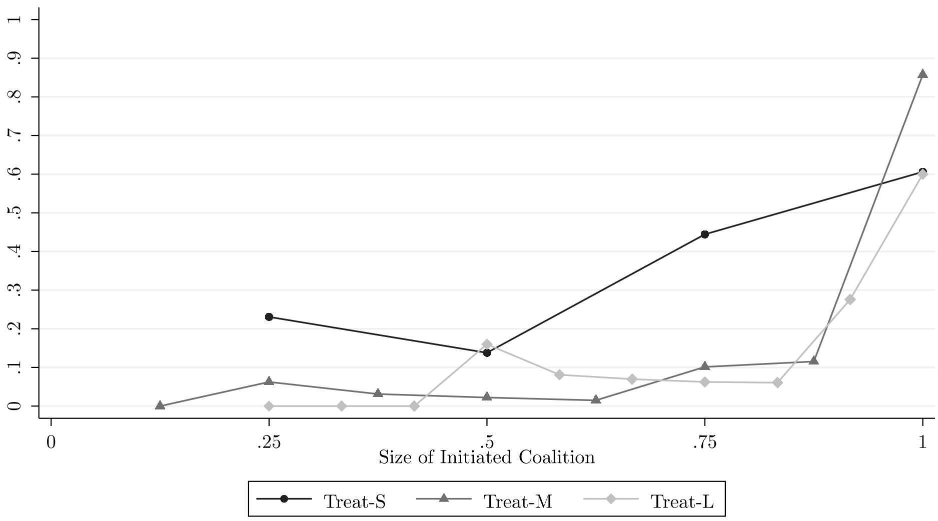
[17] As Fig. 1 indicates, the difference in the implementation rate of other coalitions comes from non-grand, non-profitable coalitions (i.e., with a relative size of less than 50 % of the group).

A. Frequency of Initiated Coalition Sizes



Notes: Distribution of the relative size of the initiated coalition (i.e., the share of participants) across all rounds (conditional on initiating an institution). All data are pooled on the treatment level.

B. Implementation Rate by Coalition Size



Notes: Average implementation rate of an institution, depending on the relative size of the initiated coalition across all rounds. All data are pooled on the treatment level.

Fig. 1. Institution Formation and Group Size.

affected by group size, the preference threshold for rejecting profitable, non-grand coalitions increases with group size, because relatively larger almost-grand coalitions are possible (i.e., which are more profitable and feature lower inequality). In this case, we would expect participants to vote more frequently in favor of the largest non-grand coalition in *Treat-M* and *Treat-L* than in *Treat-S*. This is not the case: the share of individuals voting in favor of these coalitions actually decreases from 76 % in *Treat-S* to 70 % in *Treat-M*, while it increases to 90 % in *Treat-L*, but the difference is not statistically significant ($p = .150$). While this may indicate that aversion towards disadvantageous inequality is not a key driver of players' voting behavior, it is also consistent with social preferences being sufficiently strong such that the (increasing) preference threshold does not bind.

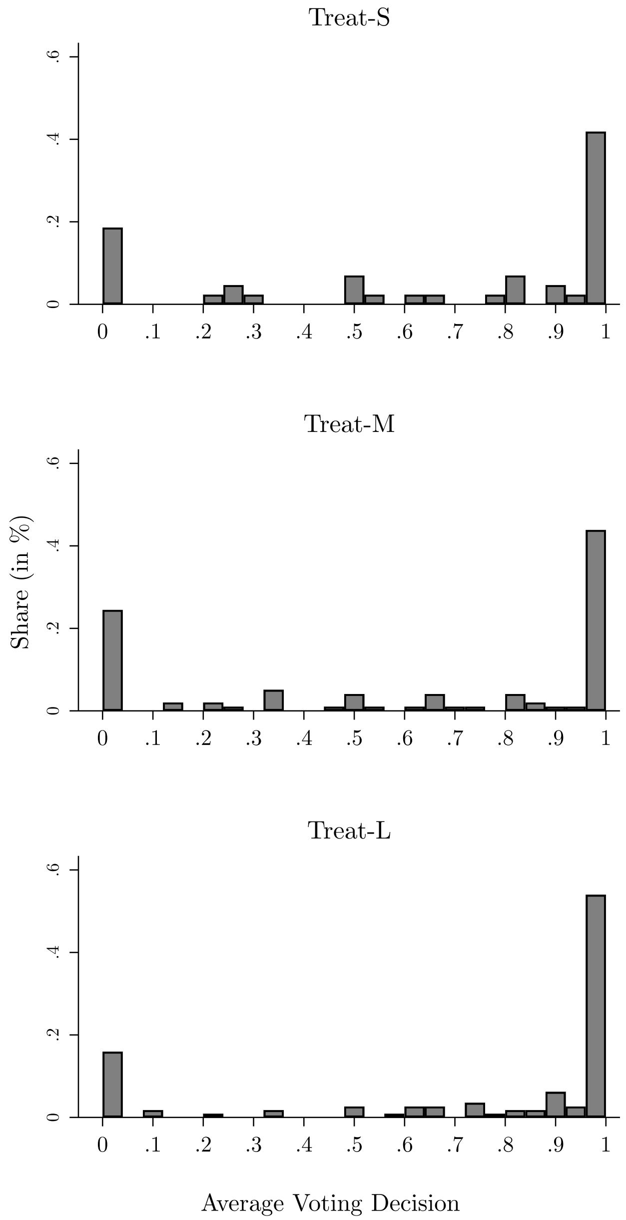


Fig. 2. Distribution of Preferences over Almost-Grand Coalitions. *Notes:* The Figure shows the histogram of average individual voting decisions on almost-grand coalitions (across all rounds) of all players who were participants in at least one initiated almost-grand coalition. All data are pooled on the treatment level.

Second, preferences themselves might be sensitive to group size. Fig. 2 plots the distribution of each average individual voting decision on almost-grand coalitions, i.e., a measure for individual preferences over almost-grand coalitions (and the associated inequality), across treatments. If group size was to affect preferences, we would expect the distribution of this preference measure to depend on group size. This is not the case. Fig. 2 highlights similarly polarized preferences, with most participants willing to implement almost-grand coalitions rather than having no institution at all. However, around 20 % of subjects are strict naysayers and are always voting against the implementation of almost-grand coalitions.[18] The share of naysayers and the share of subjects who always

[18] Importantly, when a grand coalition is initiated, naysayers almost always vote in favor of implementation (96 % in *Treat-S*, 99 % in *Treat-M*, and 100 % in *Treat-L*). In Appendix C, we show that naysayers are less likely to overestimate the number of participants, suggesting that they are aware that a grand coalition is unlikely to be initiated but are still voting against the implementation of almost-grand coalitions. A simple calibration exercise in Appendix C also shows that the share of naysayers is broadly comparable to the share of subjects in previous experiments whose *measured* preferences towards inequality would lead them to reject non-grand coalitions. Taken together, this corroborates our interpretation of their

vote in favor of almost-grand coalitions are not statistically different across group sizes ($p \geq 0.331$).[19] Third, group size naturally increases the likelihood of having at least one naysayer per group who can block almost-grand coalitions. Indeed, the share of groups with at least one naysayer increases from 42 % in *Treat-S* to 77 % in *Treat-M* ($p = .111$) and even 90 % in *Treat-L* ($p = .031$).[20] The variation in these shares also explains the variation in implementation rates of almost-grand coalitions across and within treatments (Spearman's $\rho \leq -0.5609$ with $p \leq 0.046$). We therefore attribute the lower implementation rate of almost-grand coalitions to the unanimity rule and the associated increase in the likelihood of disagreement in larger groups and not a change in preferences.[21]

Finally, consider that fewer grand coalitions are initiated in larger groups. The data suggest that participants very often continue to participate in the next round even when an almost-grand coalition was rejected (86 % in *Treat-S*, 89 % in *Treat-M*, and 92 % in *Treat-L*). The lower frequency of initiated grand coalitions is thus due to nonparticipants not changing their mind in favor of participation in the next round when an almost-grand coalition is rejected. Miscoordination could explain this, with group size making it more difficult for multiple nonparticipants to collectively change their mind about commitments.[22] Miscoordination, however, would imply that the likelihood of any individual nonparticipant changing her mind in favor of participation should be comparable between small and large groups. This is not the case. When almost-grand coalitions fail to be implemented, the likelihood of a nonparticipant changing her mind and participating in the following round decreases from 45 % in *Treat-S* to 28 % in *Treat-M* and *Treat-L*. This finding is striking, considering that the payoff from a grand coalition exceeds the payoff from no institution.[23] Consequently, increasing group size beyond four players not only severely hinders the formation of almost-grand coalitions through the unanimity rule, but also prevents the rejection of almost-grand coalitions from resolving the heterogeneity in the willingness to participate in favor of grand coalitions.

5. Conclusion

We study the effect of group size on the formation of institutions that enforce cooperation commitments and require unanimity to form. To isolate the causal effect of group size, we de-couple group size from the profitability of an institution and keep profitability constant across treatments. We find that previous results, highlighting the successful formation of grand and almost-grand coalitions that boost cooperation (e.g., Kosfeld et al., 2009), do not generalize to group sizes beyond four players. Grand coalitions are implemented but are almost never initiated in larger groups, while almost-grand coalitions are initiated but almost never implemented. The former is caused by nonparticipants who are more reluctant to change their mind about commitments, the latter stems from the unanimity rule coupled with similarly polarized preferences over the implementation of almost-grand coalitions. We also find that cooperation is lower in larger groups than the lower implementation rate would predict, suggesting that group size influences cooperation through multiple channels in this setting.

While simply requiring a majority decision seems like a solution at first glance, the lack of authority to enforce rules lies at the heart of many collective action problems. Our results therefore suggest that institutions that, due to the lack of higher authority, rely on the unanimity rule are unlikely to succeed when group size becomes large. For example, prominent successful institutions implemented through non-grand coalitions in the context of climate clubs, such as the Paris Climate Agreement and the European Emissions Trading System, either drop formal enforcement mechanisms in favor of voluntary targets (Paris Climate Agreement) or rely on majority voting procedures backed by existing institutions (European Emissions Trading System). Alternatively, one may want to relax the commitment level and let coalitions endogenously choose a minimum cooperation level. However, Dannenberg et al. (2014) show that this does not improve on an institution with full contribution in groups of ten. Future research may therefore study mechanisms for participants to reach a consensus on tolerating single or few nonparticipants and to motivate nonparticipants to change their mind (e.g., through the activation of social norms).

Data availability

Paper links to public OSF repository with data and code.

---

average individual voting decision on almost-grand coalitions in terms of preferences over almost-grand coalitions and the associated inequality. In Appendix C, we also analyze whether naysayers are different from other participants along socio-economic dimensions and performance in control questions, but do not find statistically significant differences.

[19] Group size does not affect other parts of the distribution either (more details in Appendix C).

[20] Given the same share of naysayers, the likelihood of having at least one naysayer in the group cannot decrease with group size. A one-sided test therefore has more statistical power to detect such a directional effect. Indeed, the equivalent one-sided Fisher's test gives $p = .082$ ($p = .026$) for the comparison between *Treat-S* and *Treat-M* (*Treat-L*).

[21] Another potential driver of individual behavior could be group-size effects on the belief about the number of other participants, in particular if participants in larger groups were systematically more likely to overestimate the number of other participants. In Appendix C, we show that this is not the case.

[22] *Treat-M* and *Treat-L* typically feature more than one nonparticipant in almost-grand coalitions.

[23] In Appendix C, we show that nonparticipants are actually more likely to underestimate the size of the initiated coalition. It is striking that this does not translate into a change of mind in the following round. Explanations could be that social pressure might decline (Olson, 1965) and moral wiggle room might increase (Dana et al., 2007) as groups become larger.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We thank three anonymous reviewers as well as seminar participants at Goethe University and Johannes-Gutenberg-Mainz University for helpful comments and suggestions. We gratefully acknowledge support from the Experiment Center at the Leibniz Institute for Financial Research SAFE. IRB approval was obtained from the joint ethics committee of Goethe University and the University of Mainz prior to conducting the experiment.

Appendix A. Experimental details

Subjects played the institution formation game as described in Section 2. Before subjects could start with the experiment, they had to answer a set of control questions. See Supplemental Appendix B for details. We piloted the experiment in August 2020 and conducted the sessions in August and September 2020 with oTree (Chen et al., 2016). 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.[24] All experimental sessions were embedded in 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 anonymized participant IDs (6 random characters) by email 24 h before a session started. Subjects were only allowed into the meeting with their names set to the participant IDs, i.e., subjects did only see the anonymous IDs of other subjects in the Zoom meeting. 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 of problems with the internet connection (see below). The experimenter then shared a link to the session on oTree and subjects had to enter their ID again. Subjects never got to know the IDs of others in the group they played the game in. 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 et al. (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).

In total, 308 subjects completed the experiment.[25] On average, the experiment lasted 64 min and subjects earned 11.63€, with every point in the experiment worth 0.60€.

Appendix B. Instructions & 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!

[24] Snowberg and Yariv (2021) and Prissé and Jorrat (2022) 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.

[25] Before the institution formation game, subjects also participated in another, unrelated task (measurement of willingness to cooperate conditionally in a 2-player-public goods game, when the other player is randomly drawn from a larger group; we do not find any statistically significant effect of the group size from which the other player is drawn on individual's willingness to cooperate conditionally). The strategy method was used to elicit all decisions in this unrelated task. Therefore, there was *no* interaction between group members. We simply re-used the same group matching in the institution formation game. Which player an individual was matched to as well as the outcome and payoff from this task was only revealed at the end of the whole experiment.

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 Fig. B.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?

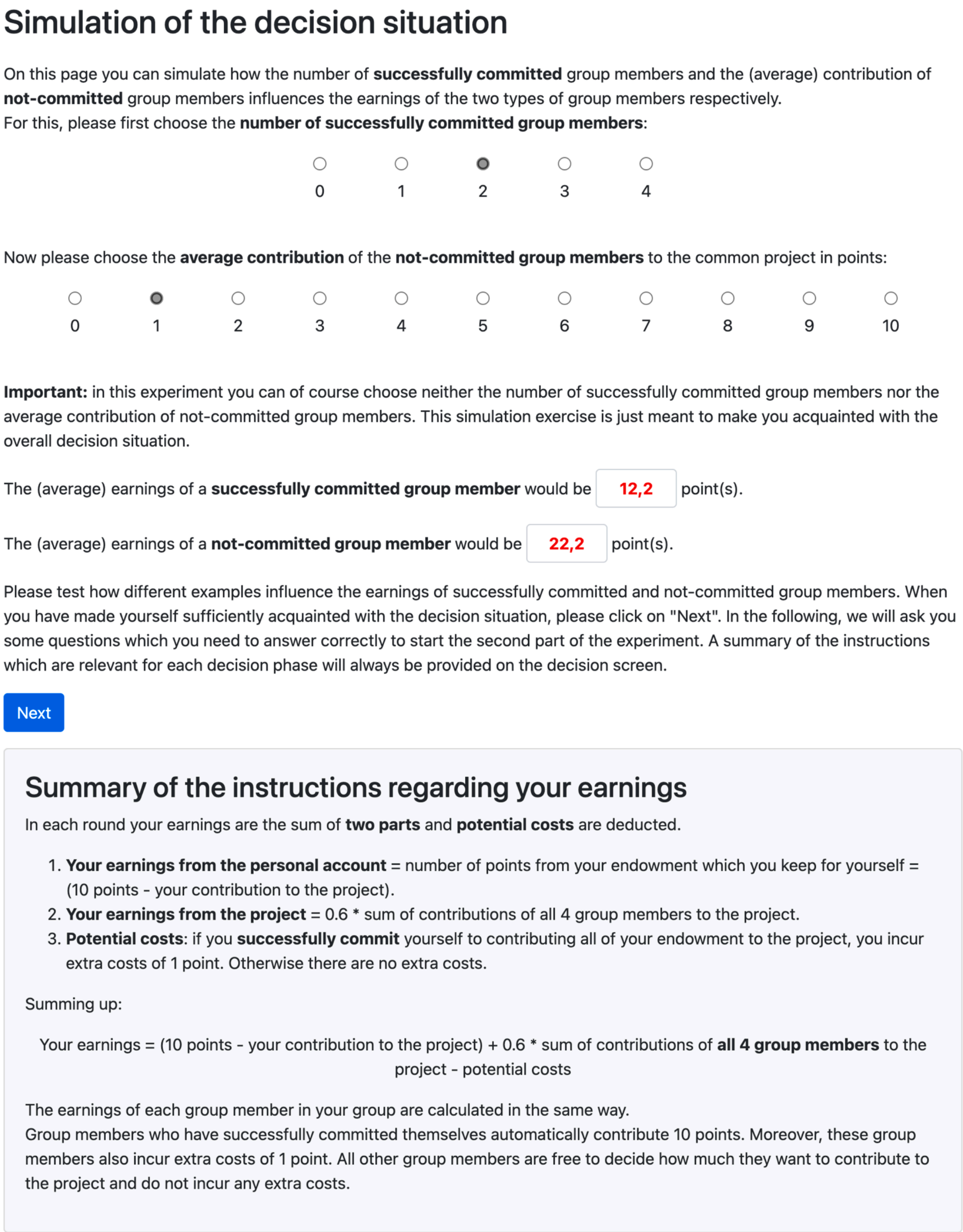
Question 3a: Suppose that nobody has committed herself to contributing all of her endowment to the project. Player A (A and E; A, E, and I) contributes **2 points**, player B (B and F; B, F, and J) contributes **4 points**, player C (C and G; C, G, and K) contributes **5 points**, and player D (D and 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 Figs. B.2 to B.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.

Fig. B.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.

Fig. B.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.

Fig. B.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.

Fig. B.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)

Fig. B.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

Fig. B.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

Fig. B.7. Fourth phase (Results).

Appendix C. Additional analyses

C.1. Group-size effect on preference distribution

Fig. 2 shows that the distribution of preferences over the implementation of almost-grand coalitions is polarized and very similar across all treatments. Statistical tests indicate that there is no evidence in favor of a group-size effect on these distributions. First, the share of subjects who express a clear preference against almost-grand coalitions (by always voting against the implementation) and the share of subjects who prefer an almost-grand coalition over no institution (by always voting in favor of the implementation) are not statistically different across group sizes at conventional significance thresholds ($p \geq 0.331$). Second, this also holds when we consider subjects who sometimes vote in favor (or against) the implementation of an almost-grand coalition: the shares of subjects who voted in favor of the implementation at most $x\% \in \{10, 20, 30\}$ or at least $y\% \in \{70, 80, 90, 100\}$ are not significantly different between *Treat-S* and any of the two larger group size treatments (cf. Table C.1; $p \geq 0.295$). The same holds for the standard deviation in the individual average voting decision on almost-grand coalitions ($p \geq 0.164$). Thus, group size does not seem to affect the distribution of individual preferences over almost-grand coalitions.

C.2. Analysis of naysayers

Table C.2 explores whether subjects who always vote against the implementation of almost-grand coalitions (naysayers) are different from people who vote at least once in favor of the implementation of an almost-grand coalition. We test whether the differences are statistically significant along a variety of dimensions (age, sex, study program, performance in the control questions), using a non-parametric Mann-Whitney-U test on the individual level and pooling all treatments to ensure sufficient statistical power.[26] We do not find any statistically significant differences ($p \geq 0.259$). However, we do find that naysayers have more pessimistic beliefs about the number of participants ($p = .001$) and are actually less likely to overestimate the number of participants in the first stage ($p = .006$). This suggests that naysayers are aware that not everyone will participate but still vote against the almost-grand coalition. Moreover, as already mentioned in Section 4, when a grand coalition is initiated, the same naysayers vote in favor of implementation (96 % in *Treat-S*, 99 % in *Treat-M*, and 100 % in *Treat-L*), corroborating our interpretation that their behavior is motivated by strong *preferences* against almost-grand coalitions.

Finally, to compare the share of naysayers in our experiment to previous findings, we compute the Fehr-Schmidt-parameter α_i which would be in line with this "naysaying-behavior". To do so, we consider coalition sizes of 75 % and compute α_i such that the utility of voting in favor of the coalition is smaller than the utility of voting against the implementation. This gives a lower bound for α_i (rejecting a relatively larger coalition requires an even higher α_i). For this, we use Eq. (E.5) from Appendix E and use the actual (average) contributions for the respective cases as proxies for beliefs about contributions of others in case of no implementation and about contributions of nonmembers in case of successful implementation. The resulting 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* respectively. Blanco et al. (2011) explicitly measure α_i in a sample of 72 students and find that roughly 18 out of 61 subjects (11 subjects are dropped from their analysis due to non-unique switching points) exhibit behavior which implies $\alpha_i \geq 1.25$ in their experiment (cf. Figure 1 in Blanco et al., 2011). This result (29.5 %) is even a bit higher than the share of naysayers among all subjects (i.e., also those who never participated in at least 1 almost-grand coalition) in our experiment (around 15–20 %), highlighting that the parameters required to rationalize naysaying behavior in our experiment are not unrealistically high.

[26] Results are identical when we use a parametric t-test. Due to the interaction within groups, individual observations are not statistically independent. Results are qualitatively similar (but more difficult to interpret) when we use a Wilcoxon signed-rank for paired samples and compare the average value for naysayers and non-naysayers within a group.

Table C.1
Share of individuals by average acceptance rate of almost-grand coalitions.

Share of Individuals (in %)	*Treat-S*	*Treat-M*	*Treat-L*
Voting in favor of almost-grand: never	18.60	24.49	15.93
Voting in favor of almost-grand: at most			
10 %	18.60	24.49	15.93
20 %	20.93	27.55	18.58
30 %	27.91	29.59	18.58
Voting in favor of almost-grand: at least			
70 %	58.14	53.06	70.80
80 %	55.81	52.04	66.37
90 %	48.84	45.92	59.29
Voting in favor of almost-grand: always	41.86	43.88	53.98

Notes: Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. Subjects who never made an implementation decision on an almost-grand coalition are excluded. All data are pooled at the treatment level.

Table C.2
Differences between naysayers and remaining participants.

	Naysayers	Others	*p*
Age (in years)	24.56	24.96	0.92
Male (share in %)	38.78	47.50	0.35
Economics or Business Major (share in %)	34.00	37.25	0.80
Incorrect answers per control question	0.68	0.63	0.26
Belief about Share of Participants (in %)	55.00	67.93	0.01
Likelihood to Overestimate Number of Participants	0.22	0.31	0.01

Notes: Comparison of age, sex, and study program (whether subjects pursue an Economics or Business program) as well as performance in control questions (avg. number of incorrect answers per control question) between naysayers (subjects who always vote against the implementation of an almost-grand coalition) and other subjects who participated in at least one almost-grand coalition. Observations are pooled across all treatments. p-value from a non-parametric Mann-Whitney-U test.

Table C.3
Beliefs.

	Treat-S	*Treat-M*	*Treat-L*
Belief about Share of Participants (in %)	68.77	59.82	62.23
Likelihood to Overestimate Share of Participants			
By Participants	0.31	0.32	0.36
By Nonparticipants	0.14	0.20	0.16
Likelihood to Underestimate Share of Participants			
By Participants	0.17	0.26	0.39
By Nonparticipants	0.39	0.57	0.68

Notes: Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data are pooled on the treatment level. Beliefs are elicited about the number of participants among other players.

C.3. Group-size effect on beliefs

Table C.3 summarizes players' belief about the share of participants. Importantly, beliefs were elicited about the number of participants among other players (instead of total participants) to improve ease-of-use. We therefore transform the elicited belief into the belief about the total number of participants (i.e., for participants, this also includes themselves).

On average, beliefs are are a bit lower in *Treat-M* and *Treat-L* compared to *Treat-S*, but these differences are not statistically significant ($p \geq 0.158$). However, this hides heterogeneity across all treatments: participants (nonparticipants) have lower (higher) beliefs when group size increases, but only the increase among participants is statistically significant ($p = .004$ for the comparison between *Treat-S* and *Treat-M* and $p = .003$ for the comparison between *Treat-S* and *Treat-L*).

Remember that the size of the initiated coalitions also decreases in larger groups. It is therefore interesting to analyze how frequently players were incorrect and had to update their beliefs. Importantly, if participants in larger groups were systematically more likely to overestimate the share of participants (i.e., they would be expecting larger initiated coalitions), the resulting downwards belief updating (i.e., the disappointment) could explain the lower willingness to implement (almost-grand) coalitions. Conversely, suppose that nonparticipants were too pessimistic about the share of participants (i.e., nonparticipants underestimate the size of the initiated coalition). This could motivate them to change their mind in the following round. To explain our results of nonparticipants not changing their mind in larger groups, we would thus expect that nonparticipants are systematically more pessimistic in *Treat-S*, prompting them more frequently to change their mind.

First, consider participants. Participants are a bit more likely to overestimate the size of the initiated coalition (0.31 in *Treat-S* compared to 0.32 in *Treat-M* and 0.36 in *Treat-L*), but these differences are not statistically significant ($p \geq 0.727$). Moreover, within all treatments, being a naysayer, i.e., a participant who actually blocks the implementation of almost-grand coalitions, is negatively correlated with the average likelihood to overestimate the number of other participants (Spearman's $\rho \leq -0.0772$ with $p \leq 0.001$).

Second, consider nonparticipants. Nonparticipants are actually substantially more likely to underestimate the number of participants in larger groups ($p = .007$ for the comparison between *Treat-S* and *Treat-M* and $p = .005$ for the comparison between *Treat-S* and *Treat-L*). That this positive surprise about the number of participants does not translate into a higher willingness to participate is surprising.

C.4. Robustness of results to learning

Since the institution formation game is a complex game and players might learn over time, results might change when we look at later rounds. However, all results are qualitatively similar when we only consider rounds 11–20 for our analyses.

Table C.4
Institution formation & contributions (Rounds 11–20).

	Treat-S	*Treat-M*	*Treat-L*
A. Implementation Success			
Formation Rate (in %)	42.86	17.33	14.55
Initiation Rate (in %)	97.86	98.00	100.00
Implementation Rate (in %)	43.80	17.69	14.55
Relative Size of Implemented Coalition (in %)	84.17	93.27	81.77
B. Public Good Contributions			
Contribution per Player	6.10	2.76	2.07

Notes: Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data are pooled on the treatment level for rounds 11–20 only.

Table C.5
Contributions by implementation outcome (Rounds 11–20).

	Treat-S	*Treat-M*	*Treat-L*
A. Successful Implementation			
Frequency (in %)	42.86	17.33	14.55
Contribution per Player	9.33	9.39	8.74
From Members	8.42	9.33	8.18
From Nonmembers	0.91	0.07	0.57
B. Failed Implementation			
Frequency (in %)	55.00	80.67	85.45
Contribution per Player	3.57	1.39	0.94
From Participants	2.18	1.02	0.66
From Nonparticipants	1.39	0.37	0.28

Notes: Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data are pooled on the treatment level for rounds 11–20 only. We omit the case when no institution was initiated, since an institution was almost always initiated across group sizes (cf. Table C.4). Contributions are normalized by dividing the average total contribution from all players, members, participants, etc. by the number of total players in the respective treatment.

Table C.4 replicates Table 1 for rounds 11–20. Both formation rate and implementation rate are again significantly different when group size increases: implementation rates decrease from 0.44 in *Treat-S* to 0.18 in *Treat-M* ($p = .021$) and 0.15 in *Treat-L* ($p = .021$). Also in rounds 11–20, contributions are low, in particular in larger groups ($p = .004$ for the difference between *Treat-S* and *Treat-M* and 0.001 for the difference between *Treat-S* and *Treat-L*).

Similarly, Table C.5 highlights that contributions per player are lower than what the lower implementation rate would predict. While Panel A shows that contributions per player in case of a successful implementation are similar across all treatments ($p \geq 0.538$), Panel B documents that contributions per player are lower in larger groups in case of a failed implementation ($p = .030$ for the difference between *Treat-S* and *Treat-M*, $p = .011$ for the difference between *Treat-S* and *Treat-L*). Importantly, this again holds for both participants and nonparticipants.

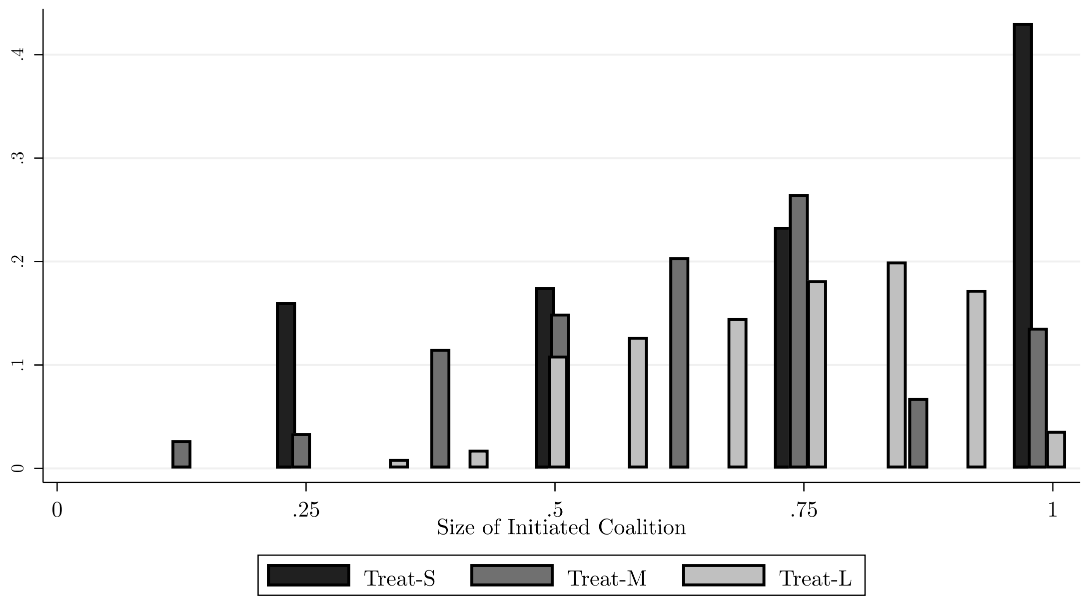
Second, consider that these aggregate effects are driven by a lower implementation rate of almost-grand coalitions, coupled with a lower frequency with which grand coalitions are initiated. Panel A and Panel B in Fig. C.1 replicate Fig. 1 for rounds 11–20.

Fig. C.1 again shows a marked difference between small and larger groups in rounds 11–20 for both outcomes. Coalitions with three out of four subjects are more frequently implemented than coalitions 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 coalition 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 %-coalition remains statistically significant ($p = .095$) at conventional levels ($p = .142$ for the difference in implementation rate of 75 % coalitions between *Treat-S* and *Treat-M*).[27]

Panel A in Fig. C.1 shows that also in rounds 11–20 grand coalitions are initiated significantly less frequently in larger groups ($p = .012$ for the difference between *Treat-S* and *Treat-M* and $p = .006$ for the difference between *Treat-S* and *Treat-L*). Table C.6 replicates Table 3 and replicates these insights for the different (aggregated) coalition types, i.e., grand coalitions, almost-grand coalitions, and other coalitions.

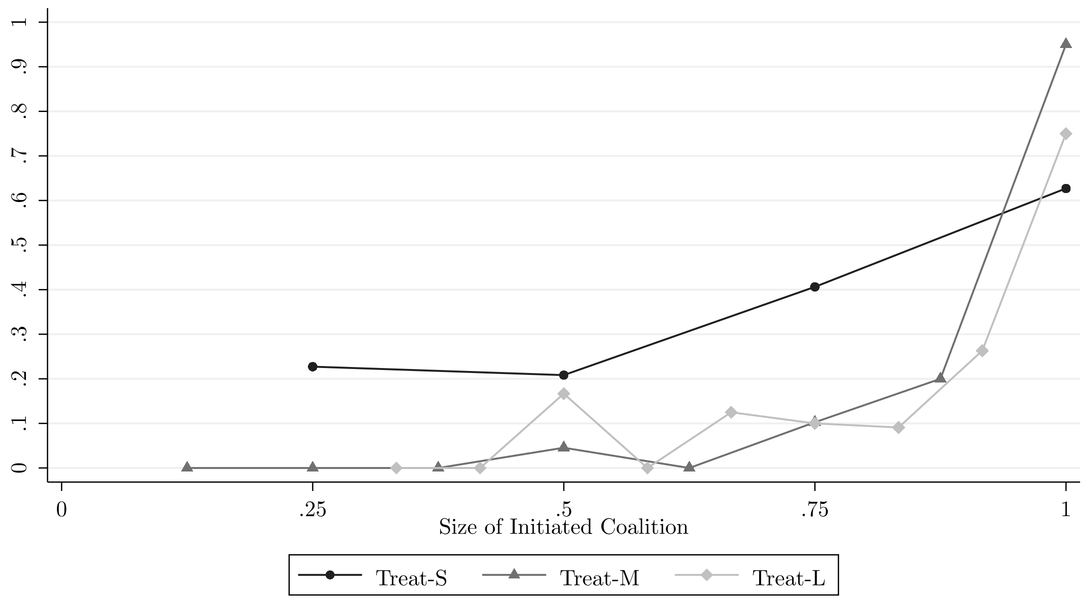
[27] To be specific, for the grand coalition, $N_G = 10$ for *Treat-S*, $N_G = 3$ for *Treat-M*, and $N_G = 2$ for *Treat-L*. For the largest non-grand coalition: $N_G = 9$ for *Treat-S*, $N_G = 5$ for *Treat-M*, and $N_G = 7$ for *Treat-L*.

A. Frequency of Initiated Coalition Sizes



Notes: Distribution of the relative size of initiated coalition (i.e., the share of participants) in rounds 11-20 (conditional on initiating an institution). All data are pooled on the treatment level.

B. Implementation Rate by Coalition Size



Notes: Average implementation rates of institutions depending on the relative size of the initiated coalition in rounds 11-20. All data are pooled on the treatment level.

Fig. C.1. Institution Formation and Group Size (Rounds 11–20).

Table C.6
Institution formation by coalition size (Rounds 11–20).

	Treat-S	*Treat-M*	*Treat-L*
A. Share of Initiated Coalitions (in %)			
Grand Coalition	43.07	13.61	3.64
Almost-Grand Coalition	23.36	33.33	55.45
Other	33.58	53.06	40.91
B. Conditional Implementation Rate (in %)			
Grand Coalition	62.71	95.00	75.00
Almost-Grand Coalition	40.63	12.24	14.75
Other	2.17	1.28	8.89

Notes: A grand coalition requires everyone to participate, i.e., $s/n = 1$. An almost-grand coalition refers to $s/n \in [0.75, 1)$, with other referring to $s/n \in [0, 0.75)$. Group size (n) is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players. All data are pooled on the treatment level for rounds 11–20 only.

Now consider the individual channels. First, consider that the threshold for rejecting non-grand coalitions increases with group size. Also in rounds 11–20, participants do not vote more frequently in favor of the largest non-grand coalition in *Treat-L* or *Treat-M* than in *Treat-S* ($p \geq 0.284$).

Second, consider heterogeneous preferences over almost-grand coalitions. Fig. C.2 replicates Fig. 2 for rounds 11–20. It clearly shows that group size does not affect revealed preferences over almost-grand coalitions. Similar to the main analysis, the great majority of participants in an almost-grand coalition consistently vote in favor of implementation. Again, roughly 20 % of participants, however, always vote against the implementation of such coalitions. Any difference in the share of these subjects is not significantly different across group sizes ($p \geq 0.202$). Moreover, also the share participants who voted at most $x \in \{10, 20, 30\}$ or at least $y \in \{70, 80, 90, 100\}$ in favor of implementation of an almost-grand coalition is not significantly different between *Treat-S* and any of the two larger group size treatments ($p \geq 0.202$). Consequently, group size does not seem to affect the distribution of preferences.

Third, we also document for rounds 11–20 that the share of groups with at least one subject who always votes against the implementation of almost-grand coalitions 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* ($p = .066$ for the one-sided Fisher's exact test with more power, $p = .131$ for the two-sided Fisher's exact test; $p = .046$ for a simple χ^2-test.). Similarly, having at least one such naysayer also explains the variation in the implementation rates of almost-grand coalitions in rounds 11–20 (Spearman's ρ = -0.7771 with $p = .000$).

Finally, consider the finding that the lower frequency of initiated grand coalitions is driven by a lower willingness of nonparticipants to change their mind. Also in rounds 11–20, nonparticipants are less willing to participate in the following round when an almost-grand coalition was rejected (87 % in *Treat-S*, 89 % in *Treat-M*, and 94 % in *Treat-L*). The likelihood that a nonparticipant participates in the following round decreases from 47.62 % in *Treat-S* to 22.22 % in *Treat-M* and 26.04 % in *Treat-L*.

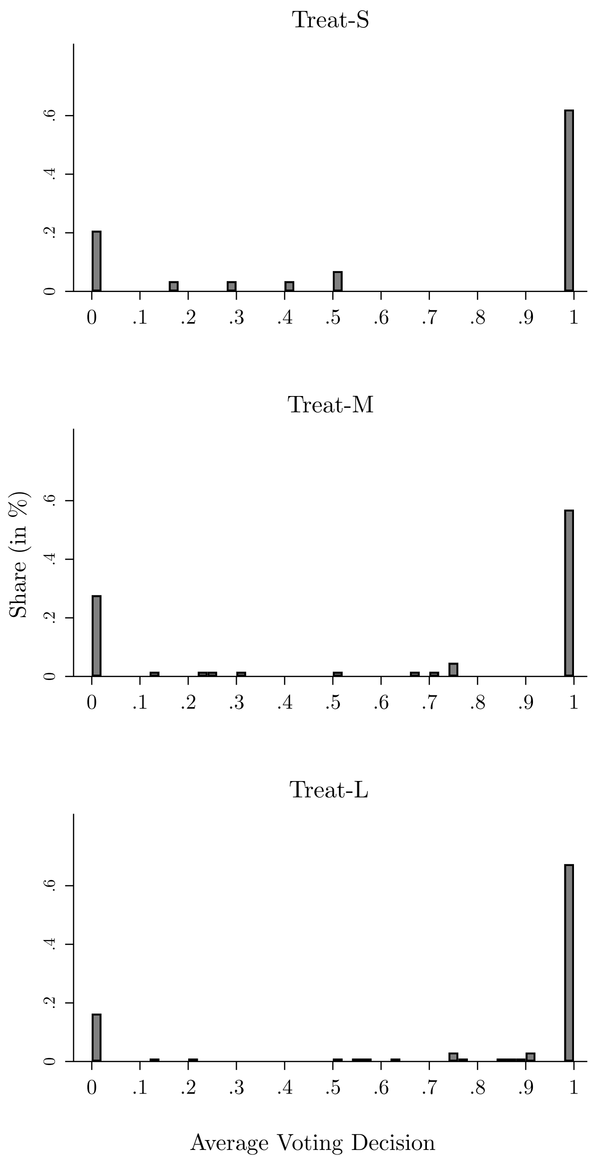


Fig. C.2. Distribution of Preferences over Almost-Grand Coalitions (Rounds 11–20). *Notes:* Distribution of preferences over the implementation of almost-grand coalitions, i.e., the average voting decision on almost-grand coalitions in rounds 11–20, among players who were participants in at least one initiated almost-grand coalition. All data are pooled on the treatment level.

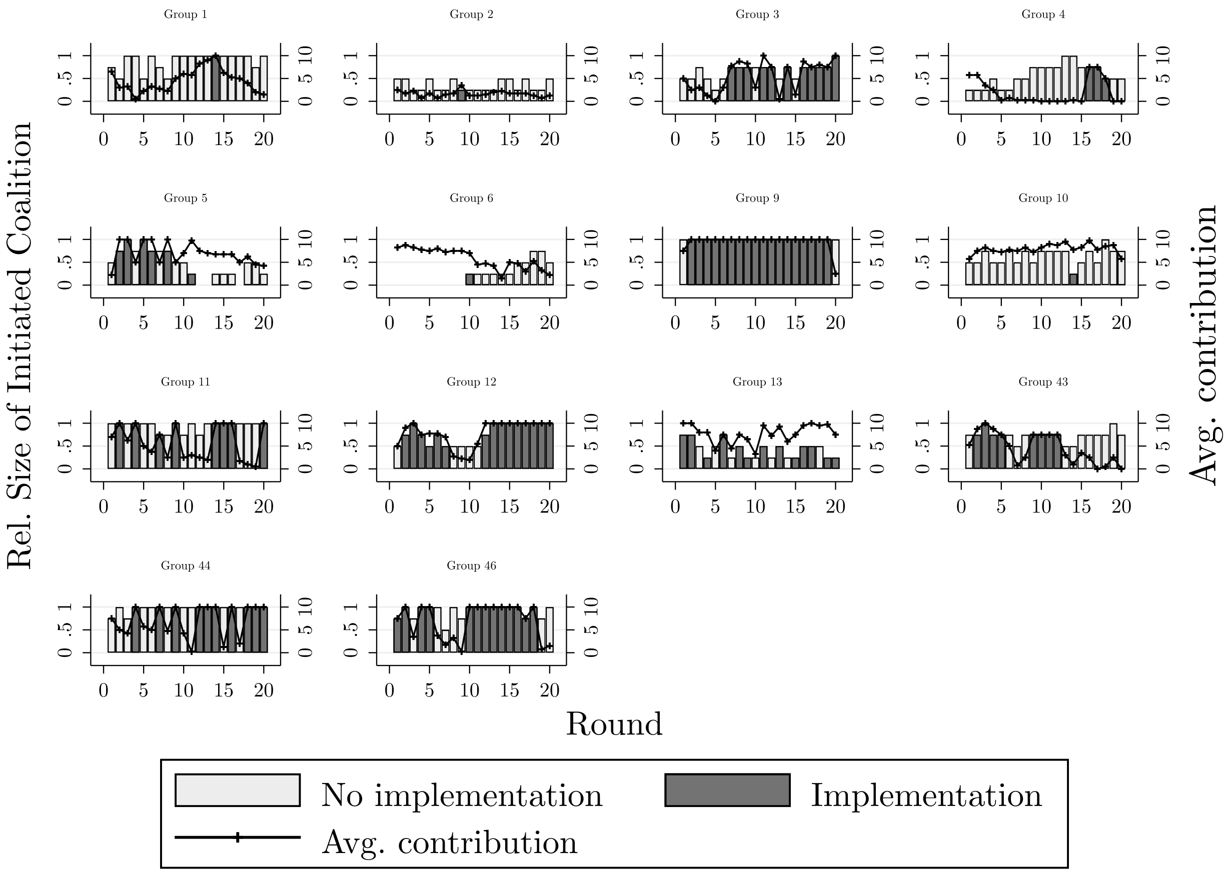


Fig. D.1. Behavior of Groups in *Treat-S*. *Notes:* Behavior of each group in *Treat-S* for all rounds, including the relative size of the initiated coalition, the final implementation decision, and average contributions per player. In case no institution was initiated, the relative size of the initiated coalition is zero. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players.

Appendix D. Behavior in every group

Figs. D.1–D.3 give an overview of the key outcomes for every group in every round of our experiment.

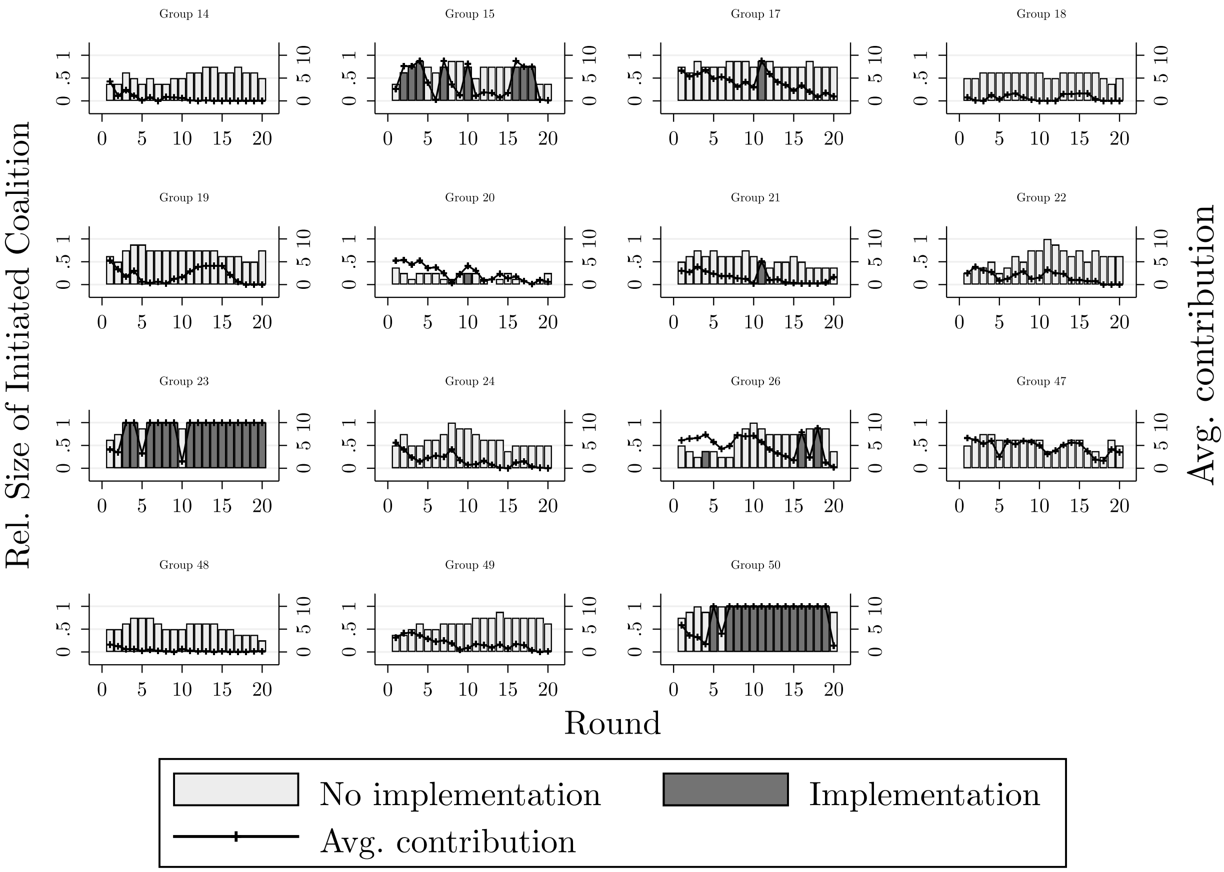


Fig. D.2. Behavior of Groups in *Treat-M. Notes:* Behavior of each group in *Treat-S* for all rounds, including the relative size of the initiated coalition, the final implementation decision, and average contributions per player. In case no institution was initiated, the relative size of the initiated coalition is zero. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players.

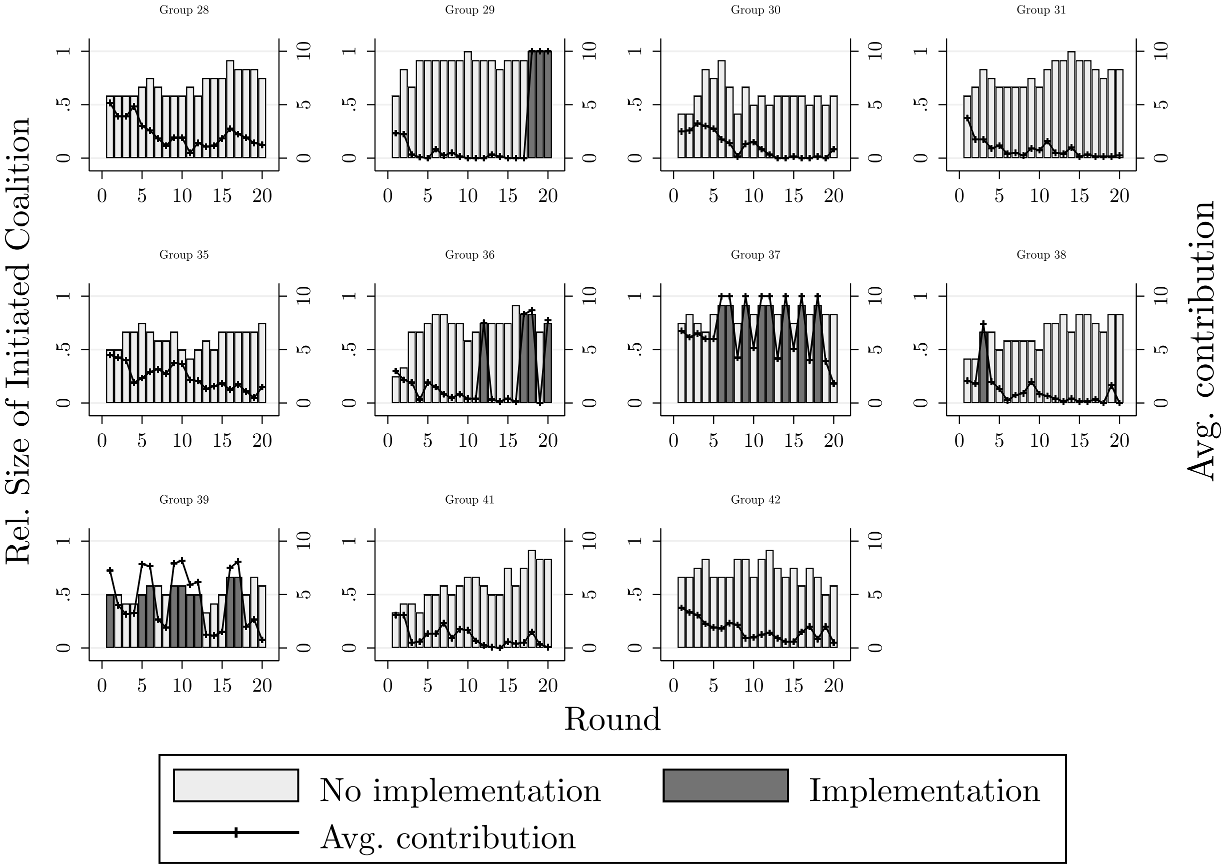


Fig. D.3. Behavior of Groups in *Treat-L*. *Notes:* Behavior of each group in *Treat-S* for all rounds, including the relative size of the initiated coalition, the final implementation decision, and average contributions per player. In case no institution was initiated, the relative size of the initiated coalition is zero. Group size is 4 (*Treat-S*), 8 (*Treat-M*), or 12 (*Treat-L*) players.

Appendix E. Institution formation game

In this section, we describe the pure-strategy equilibria of the institution formation game and how they depend on group size. As laid out in Section 2, the game 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 have homogeneous money-maximizing preferences and when (some) players have preferences against the implementation of non-grand coalitions. For the latter type of preferences, we follow Kosfeld et al. (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 et al. (2009) (with the exception of the group-size effects).

E.1. Money-maximizing preferences

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

$$\frac{\delta}{n} s w - c > w. \tag{E.1}$$

Let s^* denote the smallest integer s satisfying this trade-off, characterizing the minimal coalition that 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 (E.1) 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 formed 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 et al. (2009), taking into account the modified payoff structure for the definition of s^* as derived in Eq. (E.1). From this, it follows that the grand coalition 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 coalition 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 a grand coalition 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 coalition. 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 coalition in the second stage. The threshold s^* thus characterizes the unique strict institutional equilibrium.

Rearranging condition (E.1) (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]

$$\frac{s^*}{n} = \frac{w + c}{\delta w} \tag{E.2}$$

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 a coalition, 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 a coalition. The relative size of this coalition is independent of group size, up to rounding to the next integer.*

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

$$u_i = \begin{cases} \frac{\delta}{n} s w - c & \text{if } i \in S \\ w + \frac{\delta}{n} s w & \text{if } i \notin S \end{cases} \tag{E.3}$$

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

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 a coalition of size s, with $U_i = u_i$.

Suppose that $s > s^*$ and that the coalition is implemented. Further, suppose that participant i deviates. Since the remaining $s-1$ participants still implement the coalition 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 coalition 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 coalition, 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 coalition 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 a coalition. Participants anticipating this payoff inequality might object to it and vote against the implementation of a coalition of size $s < n$. The grand coalition with $s = n$, however, does not generate any payoff inequality. Sufficiently strong social preferences thus favor the grand coalition 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:

$$U_i = u_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} max\left\{(u_j - u_i), 0\right\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} max\left\{(u_i - u_j), 0\right\}. \tag{E.4}$$

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 a coalition of size s such that the coalition is successfully implemented:

$$\frac{\delta}{n} s w - c - \frac{\alpha_i}{n-1}(n-s)(w+c) > w. \tag{E.5}$$

Consider the grand coalition $s = n$. Suppose that at least two players have sufficiently high α_i such that condition (E.5) is violated for $s = n-1$. Regardless of which player deviated from the grand coalition in the first place, at least one player will therefore object to the implementation of a coalition with $s = n-1$. That is, the grand coalition is a strict equilibrium.[29] If this condition is even violated for $n-1$ players, no group of players is willing to implement any coalition smaller than n, as the left-hand side of (E.5) increases with s. The grand coalition 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 coalition becomes the unique strict equilibrium. Proposition 3 summarizes this.

Proposition 3. *Suppose that players have social preferences as specified in Eq. (E.4) and let $C = \left\{ i | \beta_i > 1 - \frac{\delta}{n} \right\}$. The grand coalition is a strict institutional equilibrium if and only if there exist at least two players with $\alpha_i > \tilde{\alpha}$, where*

$$\tilde{\alpha} = \begin{cases} \frac{(n-1)\left(w\left(\frac{\delta}{n}(n-1)-1\right)-c\right)}{w+c} & \text{if } C = \emptyset, \\ \frac{(n-1)(w(\delta-1)-c)}{c} & \text{if } C \neq \emptyset. \end{cases} \tag{E.6}$$

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

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

1. A coalition 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 coalition S is called *pivotal* if S is profitable but $S - \{i\}$ is not.

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

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

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

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

Second, suppose that the grand coalition is implemented in a strict subgame perfect equilibrium. We show that every participant $i \in S$ is pivotal for this coalition. For this, suppose, on the contrary, that participant $i \in S$ is not pivotal, so coalition $N - \{i\}$ 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 a coalition of size s based on Eq. (E.4). 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

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

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

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

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 coalition 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 coalition 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. Consequently, the remaining $n - 1$ participants each earn

$$U_i^p(n-1) = \frac{\delta}{n}(n-1)w - c - \frac{\alpha_i}{n-1}(w+c) \tag{E.7}$$

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

$$\alpha_i > \frac{(n-1)\left(w\left(\frac{\delta}{n}(n-1) - 1\right) - c\right)}{w+c} \equiv \tilde{\alpha}_1 \tag{E.8}$$

for at least two players. Thereby, every participant of the grand coalition 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:

$$U_i^p(n-1) = \frac{\delta}{n} nw - c - \frac{\alpha_i}{n-1} c < w \tag{E.9}$$

$$\alpha_i > \frac{(n-1)(w(\delta - 1) - c)}{c} \equiv \tilde{\alpha}_2 \tag{E.10}$$

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 coalition of size s is smaller than the utility without an institution (w by assumption).

Suppose at first that $C = \emptyset$. In this case, nonmembers do not contribute to the public good and a member's payoff strictly increases in s. Moreover, inequality also strictly decreases in s. It follows that the utility of a member of any coalition with $s \leq n-1$ is maximized for $s = n-1$. If $\alpha_i > \tilde{\alpha}_1$ for $n-1$ players, clearly any coalition with $s \leq n-1$ will always be rejected by at least a single participant. The grand coalition 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 coalition with $s \leq n-1$ is maximized for $s = n-1$ and nonparticipating player $j \in C$. Clearly, the monetary payoff of any coalition 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 w. It follows that if $n-1$ players are sufficiently averse to inequality to reject this "best-case" scenario implied by Eq. (E.9), i.e., $\alpha_i > \tilde{\alpha}_2$, any coalition with $s < n$ will also be rejected. □

Finally, consider the effect of group size if players have social preferences. In the main text, we distinguish between three possible channels: First, social preferences themselves might play a weaker role when group size increases (Schumacher et al., 2017; Alós-Ferrer et al., 2022). Clearly, this could affect whether there are at least two players with $\alpha_i > \tilde{\alpha}$. Second, while the payoffs of members and nonmembers (and thus payoff inequality between them) in a coalition of the same relative size do not depend on group size (cf. Proposition 2), larger group sizes allow for larger non-grand coalitions which feature lower (average) inequality and higher profitability. This increases the threshold $\tilde{\alpha}$. 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 coalition 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.

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 of 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 coalition. The effect of group size on institution formation then depends on the interplay of the three effects described above.

References

Agrawal, A., Goyal, S., 2001. Group size and collective action: third-party monitoring in common-pool resources. Comp. Polit Stud. 34 (1), 63–93. https://doi.org/10.1177/0010414001034001003

Agrawal, A., Yadama, G., 1997. How do local institutions mediate market and population pressures on resources? forest panchayats in kumaon, India. Dev. Change 28 (3), 435–465. https://doi.org/10.1111/1467-7660.00050

Alós-Ferrer, C., García-Segarra, J., Ritschel, A., 2022. Generous with individuals and selfish to the masses. Nat. Hum. Behav. 6, 88–96. https://doi.org/10.1038/s41562-021-01170-0

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

Blanco, M., Engelmann, D., Normann, H.T., 2011. A within-subject analysis of other-regarding preferences. Games Econ. Behav. 72 (2), 321–338. https://doi.org/10.1016/j.geb.2010.09.008

Chen, D.L., Schonger, M., Wickens, C., 2016. oTree—an open-source platform for laboratory, online, and field experiments. J. Behav. Exp. Finance 9, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

Dana, J., Weber, R.A., Kuang, J.X., 2007. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Econ. Theory 33 (1), 67–80. https://doi.org/10.1007/s00199-006-0153-z

Dannenberg, A., Gallier, C., 2020. The choice of institutions to solve cooperation problems: a survey of experimental research. Exp. Econ. 23 (3), 716–749. https://doi.org/10.1007/s10683-019-09629-8

Dannenberg, A., Lange, A., Sturm, B., 2014. Participation and commitment in voluntary coalitions to provide public goods. Economica 81 (322), 257–275. https://doi.org/10.1111/ecca.12073

D'Aspremont, C., Jacquemin, A., Gabszewicz, J.J., Weymark, J.A., 1983. On the stability of collusive price leadership. Canadian J. Econ. 16 (1), 17. https://doi.org/10.2307/134972

Detemple, J., Kosfeld, M., 2025. Replication package for "The detrimental effect of group size on institution formation". OSF. October 14. Available at https://osf.io/ck2eq.

Ertan, A., Page, T., Putterman, L., 2009. Who to punish? Individual decisions and majority rule in mitigating the free rider problem. Eur. Econ. Rev. 53 (5), 495–511. https://doi.org/10.1016/j.euroecorev.2008.09.007

Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114 (3), 817–868. https://doi.org/10.1162/003355399556151

Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125. https://doi.org/10.1007/s40881-015-0004-4

Gürerk, Ö., Irlenbusch, B., Rockenbach, B., 2006. The competitive advantage of sanctioning institutions. Science 312 (5770), 108–111. https://doi.org/10.1126/science.1123633

Isaac, R.M., Walker, J.M., 1988. Group size effects in public goods provision: the voluntary contributions mechanism. Q. J. Econ. 103 (1), 179. https://doi.org/10.2307/1882648

Isaac, R.M., Walker, J.M., Williams, A.W., 1994. Group size and the voluntary provision of public goods. J. Public Econ. 54 (1), 1–36. https://doi.org/10.1016/0047-2727(94)90068-X

Kamei, K., Putterman, L., Tyran, J.-R., 2015. State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods. Exp. Econ. 18 (1), 38–65. https://doi.org/10.1007/s10683-014-9405-0

Kamei, K., Putterman, L., Tyran, J.-R., 2023. Civic engagement, the leverage effect and the accountable state. Eur. Econ. Rev. 156, 104466. https://doi.org/10.1016/j.euroecorev.2023.104466

Kosfeld, M., Okada, A., Riedl, A., 2009. Institution formation in public goods games. Am. Econ. Rev. 99 (4), 1335–1355.

McEvoy, D.M., Murphy, J.J., Spraggon, J.M., Stranlund, J.K., 2011. The problem of maintaining compliance within stable coalitions: experimental evidence. Oxf. Econ. Pap. 63 (3), 475–498. https://doi.org/10.1093/oep/gpq023

North, D.C., 1990. Institutions, Institutional Change and Economic Performance. Cambridge University Press. 1st edition. https://doi.org/10.1017/cbo9780511808678

Nosenzo, D., Quercia, S., Sefton, M., 2015. Cooperation in small groups: the effect of group size. Exp. Econ. 18 (1), 4–14. https://doi.org/10.1007/s10683-013-9382-8

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

Okada, A., 1993. The possibility of cooperation in an n-person prisoners' dilemma with institutional arrangements. Public Choice 77 (3), 629–656. https://doi.org/10.1007/BF01047864

Olson, M., 1965. The logic of collective action: Public goods and the theory of groups. Number 124 In Harvard Economic Studies, Harvard Univ. Press, Cambridge, Mass.. 21. printing edition.

Ostrom, E., 1990. Governing the Commons: The Evolution of Institutions for Collective Action. The Political Economy of Institutions and Decisions, Cambridge University Press, Cambridge; New York.

Ostrom, E., 1999. Coping with tragedies of the commons. Ann. Rev. Political Sci. 2 (1), 493–535. https://doi.org/10.1146/annurev.polisci.2.1.493

Ostrom, E., 2005. Understanding Institutional Diversity. Princeton Paperbacks, Princeton University Press.

Poteete, A.R., Ostrom, E., 2004. Heterogeneity, group size and collective action: the role of institutions in forest management. Dev. Change 35 (3), 435–461. https://doi.org/10.1111/j.1467-7660.2004.00360.x

Prissé, B., Jorrat, D., 2022. Lab vs online experiments: no differences. J. Behav. Exp. Econ. 100, 101910. https://doi.org/10.1016/j.socec.2022.101910

Schmidt, K.M., Ockenfels, A., 2021. Focusing climate negotiations on a uniform common commitment can promote cooperation. Proc. Natl. Acad. Sci. 118 (11), e2013070118. https://doi.org/10.1073/pnas.2013070118

Schumacher, H., Kesternich, I., Kosfeld, M., Winter, J., 2017. One, two, many—insensitivity to group size in games with concentrated benefits and dispersed costs. Rev. Econ. Stud. 84 (3), 1346–1377. https://doi.org/10.1093/restud/rdw043

Snowberg, E., Yariv, L., 2021. Testing the waters: behavior across participant pools. Am. Econ. Rev. 111 (2), 687–719. https://doi.org/10.1257/aer.20181065

Steimanis, I., Struwe, N., Benda, J., Blanco, E., 2025. Reducing strategic uncertainty increases group protection in collective risk social dilemmas. Working Paper University of Innsbruck.

Sutter, M., Haigner, S., Kocher, M.G., 2010. Choosing the carrot or the stick? endogenous institutional choice in social dilemma situations. Rev. Econ. Stud. 77 (4), 1540–1566. https://doi.org/10.1111/j.1467-937X.2010.00608.x

Weimann, J., Brosig-Koch, J., Heinrich, T., Hennig-Schmidt, H., Keser, C., 2019. Public good provision by large groups – the logic of collective action revisited. Eur. Econ. Rev. 118, 348–363. https://doi.org/10.1016/j.euroecorev.2019.05.019