Endogenous Sanctioning Institutions and Migration Patterns: Experimental Evidence*

Ramón Cobo-Reyes[†], Gabriel Katz[‡], Simone Meraglia[§]

Abstract

We experimentally analyze the effect of the endogenous choice of sanctioning institutions on cooperation and migration patterns. Subjects are assigned to one of two groups, are endowed with group-specific preferences, and play a public goods game. We compare an environment in which subjects can move between groups and vote on whether to implement sanctions, to one in which only one group is exogenously endowed with sanctions. We find that the possibility of voting leads to a more efficient partition of subjects across groups, higher payoffs, lower inequality, and lower migration rates. Over time, subjects tend to vote for institutions.

Keywords: Formal Sanctions, Cooperation, Migration, Voting, Experiment.

JEL Classification Numbers: C73, C91, C92, D72, H41, H73.

Declarations of interest: none.

*We are particularly indebted to Gary Charness, Carolina Di Tella, and Dan Levin for very useful comments and discussions. We also thank seminar participants at the University of Exeter BID Brown Bag seminar, University of Copenhagen, Lund University, and IFRO Copenhagen, and participants at the 2017 Institutions and Cooperation Workshop in Exeter. The authors are thankful to the University of Exeter BID and BIP Clusters for financially supporting this project. All remaining errors are ours.

[†]Corresponding author. Department of Economics, American University of Sharjah. University City, PO Box 26666, Sharjah, UAE. E-mail: rcobo@aus.edu.

[‡]Department of Politics, University of Exeter. Amory Building, Rennes Drive, Exeter, EX4 4RJ, UK. E-mail: g.katz@exeter.ac.uk.

[§]Department of Economics, University of Exeter. Rennes Drive, Exeter, EX4 4PU, UK. E-mail: s.meraglia@exeter.ac.uk.

1 Introduction

Institutions, defined as "the humanly devised constraints that shape human interaction" (North, 1990, p. 3), play a fundamental role in societies. "Good" rules – e.g., sanctioning institutions and, more generally, the rule of law – help societies solve social dilemmas and positively affect many outcomes, such as productive effort and investments (Ostrom et al., 1992; Acemoglu et al., 2001). Different societies may, however, implement different rules and, as a consequence, experience contrasting development paths (North and Thomas, 1973; North, 1981; Jones, 2003; Acemoglu et al., 2005). These differences, in turn, can affect individuals' decisions to move between societies and, as also observed in laboratory experiments, eventually generate strong migration patterns (Sahota, 1968; Gürerk et al., 2006; Mayda, 2010).

Migration flows might entail costs for both individuals and societies. First, because of social and family ties, individuals can have preferences for their "home society" and would therefore bear a cost from moving away from home (Hill, 1987; Niedomysl and Amcoff, 2011). Second, migration can negatively affect the "sending society" – and widen the welfare gap between this society and the "hosting" one – by, for instance, depleting part of its workforce (Katseli et al., 2006; Beine et al., 2007; De Haas, 2010).[1] As an example of the relationship between institutions and migration, consider academia. Scholars move across universities that implement different sets of rules (rewards and sanctions, such as monetary incentives and tenure requirements) to govern their members' contributions to the internal teaching and research environment. Moreover, leaving aside considerations about heterogeneity in individuals' abilities, well-functioning universities are often successful in attracting scholars who, by migrating, may both bear a personal cost (e.g., moving away from home) and

[1] The cost of migration to the sending society grows larger if positive spillovers exist between workers, for instance in terms of productivity (Katseli et al., 2006). Migration flows can also have (long-run) positive effects on the sending society – e.g., through the knowledge, ideas, and financial resources brought back by returning migrants (De Haas, 2010) – as well as positive and negative effects on the hosting society as, for instance, when migration causes a congestion of locally provided public goods (Dinkelman and Schulhofer-Wohl, 2015). In this paper, we abstract from these additional effects.

impose a cost on their former organization (e.g., lower number of projects shared with former colleagues).

Provided that heterogeneity in institutions is one of the factors driving migration flows and leading to inequalities between societies, it becomes important to analyze whether societies adjust their institutions to solve these issues. In this paper, we use a laboratory experiment to investigate whether, in the presence of migration costs, the possibility of individuals choosing the institutions governing their interactions allows societies to retain (and/or attract back) their members and prosper. Our research question is therefore closely related to that in Gürerk et al. (2006), who show that, when societies cannot change their institutional setting, those implementing the "weakest" institutions become depopulated.

We study a particular set of institutions: sanctions. Recently, experimental economists have devoted considerable attention to the emergence of sanctioning institutions (Kosfeld et al., 2009) and their impact on cooperation in public good games. Informal (peer-to-peer) and formal (centralized) punishment mechanisms have been shown to significantly increase the level of cooperation (Ostrom et al., 1992; Fehr and Gächter, 2000, 2002; Gächter and Herrmann, 2009; Gürerk et al., 2009; van der Heijden et al., 2009).[2] We take this finding as a starting point for our experimental design. In our baseline treatment, we employ a setting similar to Gürerk et al. (2006). At the beginning of the experiment, we randomly allocate subjects to two groups to play a Public Goods Game (PGG). Subjects interact for 30 periods. One of the groups implements sanctioning institutions, whereas the other group has no rules governing the interactions between its members. Every period, participants decide whether to (freely) move between groups. Three main features distinguish our baseline from Gürerk et al. (2006). First, we endow subjects with group-specific incentives; that is, all else equal, subjects obtain higher payoffs in the group to which they are initially assigned ("initial group"). Second, to account for the costs endured by a society as a result of migration outflows, we make the individual return from the public good depend on the group size.

[2]For a survey of this literature, see Chaudhuri (2011).

Third, we use formal rather than informal sanctioning institutions. Implementing the formal sanction scheme is costly, and each member of the group contributes an equal and fixed amount each period to fund the institution.

We compare this baseline with our main treatment, in which subjects also have the possibility of voting at fixed intervals on the institutions – either formal and costly sanctioning institutions, or no sanctions – to be implemented in their group. We focus on endogenous institutional choice over exogenously imposed institutions mainly because (*i*) imposing institutions is not possible in democratic societies, and (*ii*) institutions are more effective when chosen by individuals rather than exogenously implemented (Sutter et al., 2010; Baldassarri and Grossman, 2011; Markussen et al., 2014).

We find that the endogenous institutional choice dramatically affects migration, individual welfare, and inequality patterns. Unlike the case in which individuals cannot vote on institutions – and despite the presence of multiple equilibria caused by the relationship between group size and individual return from the public good – groups that can choose the rules governing their interactions do not significantly diverge over time in terms of their size, composition, and contributions. On average, over the 30 periods of the game, less than 73% of subjects play in their initial group when voting is not allowed, with the group endowed with sanctioning institutions attracting most of the subjects. By contrast, when institutional choice is allowed, a statistically significant higher share (83%) of subjects play in their initial group, with the majority (62%) voting to implement sanctions by the end of the experiment. Group sizes are then more balanced when subjects can vote on institutions than in the case in which institutions cannot be chosen. Our results also show that although contributions are very similar with and without institutional choice, the possibility of self-organizing leads to higher payoffs for individuals belonging to an otherwise depopulated society, and to a more equal distribution of payoffs between groups. The individual-level analysis confirms these findings and points out subjects' propensity to move back to their initial group to cast their ballot, which further emphasizes their strong ability to self-organize efficiently.

The paper proceeds as follows. Section 2 presents a review of the literature. Section 3 describes the experimental design. Section 4 introduces a formal model and presents the main theoretical predictions and the hypotheses to be tested in the empirical analysis. Section 5 reports and discusses our findings. Section 6 concludes.

2 Literature Review

Our paper continues a stream of literature that studies both the effect of institutions on human cooperation and the evolutionary properties of sanctioning institutions.

Since the work of Yamagishi (1986) and Ostrom et al. (1992), many scholars have investigated the role played by the possibility of punishing free-riders in fostering cooperation in social dilemmas. Fehr and Gächter (2000, 2002) show individuals engage in decentralized altruistic punishment – that is, they punish defectors at a cost to themselves and in the absence of material gains – which leads to high levels of cooperation. Baldassarri and Grossman (2011) study the effectiveness of centralized forms of punishment (e.g., a police force) in sustaining cooperation by using a lab in the field experiment. The authors show centralized sanctions enhance cooperation, especially when the centralized authority is elected rather than randomly chosen.

Recently, growing attention has been devoted to the endogenous choice of different types of decentralized institutions. Ertan et al. (2009) study the adoption of different punishment technologies. The authors find individuals tend to increasingly implement sanctioning institutions over time, through which only low contributors to the public good can be punished. Sutter et al. (2010) study the endogenous choice of sanctions and rewards within a standard Voluntary Contribution Mechanism. They find individuals choose the rewarding or punishment options, as opposed to a sanction-free environment, when sufficiently effective; also, individuals tend to prefer to reward cooperation rather than punish free-riding. An effect of the endogenous choice of institutions is found: higher cooperation levels arise when individuals endogenously choose the rules governing their interactions.

Because the presence of second-order free-riding and antisocial punishment lowers the effectiveness of peer-to-peer punishment (Denant-Boemont et al., 2007; Dreber et al., 2008; Herrmann et al., 2008; Nikiforakis, 2008), a strand of this literature focuses on the emergence of centralized forms of punishment. Andreoni and Gee (2012) experimentally compare decentralized to centralized punishment and find that, in the vast majority of cases, subjects implement a centralized mechanism that crowds out the peer-to-peer mechanism. Sigmund et al. (2010) theoretically show more centralized forms of punishment (i.e., pool punishment) are more stable than peer punishment, because second-order free-riding is also punished. The authors, however, find pool punishment to be less efficient than peer punishment. Zhang et al. (2014) generally confirm this prediction by means of an experiment, even though they find peer punishment to be more stable than theoretically predicted. Traulsen et al. (2012) argue a centralized punishment scheme is preferred to a decentralized one provided individuals are punished for not contributing to the funding of centralized institutions.

Markussen et al. (2014) also focus on individuals' dynamic choice of punishment mechanisms. They experimentally analyze the conditions under which a centralized punishment scheme emerges as the preferred mechanism vis-à-vis an informal mechanism as well as a sanction-free environment. As in our setting, operating the formal sanction scheme requires individuals to pay a fixed cost. The authors find that, in line with social preferences theories, a decentralized punishment mechanism is often preferred to a centralized mechanism whenever the latter becomes sufficiently costly (but profitable). By contrast, a centralized mechanism becomes the preferred choice when deterrent and sufficiently cheap.[3] Like Sutter et al. (2010), the authors further confirm the welfare-enhancing effect of the endogenous institutional choice.

Gross et al. (2016) analyze individuals' decisions to delegate punishment power to others. They show hierarchical power structures emerge out of voluntary transfers of punishment

[3]On the advantage of centralized forms of punishment over decentralized ones, see also Guala (2012).

power to those members who are more willing to engage in decentralized punishment; this process results in sustainable cooperation. A similar finding is reported in Fehr and Williams (2018). The authors show individuals prefer to delegate punishment power to a central authority rather than to engage in peer-punishment. The popularity (and effectiveness) of the latter institution, however, significantly increases when individuals can form a normative consensus about an expected level of cooperation.

Our paper complements this literature by studying the evolution and regulatory properties of sanctioning institutions when subjects can vote with both feet and ballots.

Subjects' ability to move between groups also links our paper to the literature on the relationship between institutions and migration patterns. Gürerk et al. (2006) experimentally analyze a public goods game and show that, when confronted with a sanction-free society, a society endowed with peer-to-peer punishment succeeds in attracting all the subjects and sustaining cooperation. In a similar environment, Gürerk et al. (2014) find individuals' self-selection into the community implementing the preferred institutional framework is a key determinant of long-run cooperation. We mainly differ from Gürerk et al. (2006) in that we allow each group to endogenously choose its preferred sanctioning regime.

Our focus on the co-determination of institutions, migration, and the ability to solve social dilemmas also links this paper to Robbett (2014). The author studies a Tiebout-like environment in which (i) individuals are endowed with preferences over bundles of local taxes and public good provision, and (ii) returns from the public good increase with community size. Robbett (2014) finds that, when unable to vote on their preferred bundle, individuals tend to sort optimally but fail to achieve full efficiency due to over-segregation in communities providing a Pareto-dominated bundle. This inefficiency is overcome when voting on bundles is allowed because individuals, once sorted, choose their preferred taxes and public good provision. Our research question is related to but differs from that answered by Tiebout-like models: rather than studying the effect of "voting with feet" on individuals' optimal sorting given preferences over taxation and public goods provision, we focus on the effect of the endogenous choice of institutions on individuals' sorting given preferences for location.

3 Experimental Design and Procedures

We conducted laboratory experiments at the University of Exeter between February and October 2016. Participants were mainly students of economics, business administration, and engineering, but other disciplines were also represented. We ran two different treatments with a total of 50 sessions, 25 sessions per treatment. Each session was composed of 10 subjects; the average individual earnings were £14. All instructions can be found in Appendix A.

The experiment consists of two different activities. Upon arrival to the laboratory, subjects are randomly divided into two groups of five people each. In the first activity, each subject receives 50 tokens and participates in a standard one-shot *PGG*. The outcomes of this first activity are used as instruments in individual-level analyses accounting for covariate endogeneity (see footnotes 38, 44 and 47 in Section 5.2 and Tables A.4, A.7 and A.10 in Appendix B). At the end of the first activity, subjects receive the instructions for the second task, which consists of a *PGG* that lasts 30 periods. No information about the outcome of the first activity is released. In the second activity, we run two treatments that vary depending on whether the institutional setting is endogenously chosen.

In the ***No-Voting* treatment (*NVT*)**, subjects are randomly assigned in the first period to one of two groups – A and B – of five people each. The group to which each subject is assigned in the first period is her "initial group". Each period is composed of two (sequential) stages: a *contribution stage* and a *moving stage*. In the *contribution stage*, each subject decides how to allocate the 50 tokens she is endowed with between a group account and a private account. As Table 1 shows, a subject's return from an allocation to the group account depends on the size of the group in which she is located in a given period, and on whether this group is the subject's initial group.[4]

[4]The parameter values in Table 1 are aligned with those used in previous experimental studies of PGG (see, for instance, Charness et al., 2014).

Table 1: Individual factors multiplying the group contribution to the public good and marginal per-capita return (MCPR), by group size and membership

(1) Group Size	(2) Factor, if in initial group	(3) MPCR, if in initial group	(4) Factor, if **not** in initial group	(5) MPCR, if **not** in initial group
1	1.00	1.00	1.00	1.00
2	1.50	0.75	1.15	0.58
3	1.75	0.58	1.40	0.47
4	2.00	0.50	1.65	0.41
5	2.25	0.45	1.90	0.38
6	2.25	0.38	1.90	0.32
7	2.25	0.32	1.90	0.27
8	2.25	0.28	1.90	0.24
9	2.25	0.25	1.90	0.21
10	2.25	0.23	1.90	0.19

In particular, given group size, the return from the group account is higher for subjects located in their initial group.[5] Moreover, all else equal, a subject's return from the group account is independent of whether the other group members are located in their initial group. Table 1 also illustrates that, both for subjects located in their initial group and for those located away from their home group, the factor multiplying the contributions to the group account remains constant when the group size rises above 5.

There exist two institutional settings (rule-sets) affecting subjects' payoffs.

(a) No Sanctions. Individual i's per-period monetary payoff is

$$\pi_{i,h} = (50 - C_i) + \frac{R_h(n)}{n} \sum_{j=1}^{n} C_j, \tag{1}$$

for $i, j = 1, ..., n$, where n denotes the total number of individuals located in the group,

[5] As is clear from Table 1, we model the cost for subjects of being located outside their initial group as a reduction in the return from the public good. This modeling choice captures the fact that individuals may be less willing to contribute to the public good when located away from their home society, for instance because of some form of group identity (e.g., Charness et al., 2014). As an alternative, one could model the same cost by subtracting a fixed amount from subjects' earnings. Adopting the latter alternative would not affect our (theoretical and empirical) predictions in any major way. In particular, it would not affect the *Strong Nash Stable* equilibria described in Section 4.1.4.

$C_i \in [0, 50]$ denotes i's contribution to the group account; R_h denotes the factor multiplying the contributions to the group account for individual i whose initial group is $h = A, B$, and $\frac{R_h(n)}{n}$ is the marginal per-capita return ($MPCR$) from the group account. The values of the factor R_h as a function of n are shown in column (2) of Table 1 for the case in which i's current group is h, and in column (4) if i is not located in h; the corresponding values for the marginal per-capita return are given in columns (3) and (5) of the table.

(b) **Sanctions**. Each individual in the group has to pay a fixed fee of 15 tokens per period. In addition, each individual pays a fine equal to 80% of the amount of tokens allocated to the private account in a particular period. The fixed fee and the fine (if applicable) are deducted from subjects' monetary payoff at the end of the period. Under this rule-set, i's per-period monetary payoff is therefore

$$\pi_{i,h} = (50 - C_i)(1 - 0.8) + \frac{R_h(n)}{n}(\sum_{j=1}^{n} C_j) - 15. \tag{2}$$

In *NVT*, Group A implements *No Sanctions* and Group B implements *Sanctions* for the entire duration of the game.[6]

At the end of the *contribution stage*, subjects receive information about: (i) the average contribution to the group account in their current group, (ii) the average contribution in the other group, (iii) the rule-set implemented in their current group, and (iv) the rule-set implemented in the other group. Subjects then enter the *moving stage* and simultaneously decide whether to move from their current group.

In the ***Voting* treatment (*VT*)**, besides the *contribution* and *moving* stages, there is a *voting stage* in which subjects decide on the rule-set to be implemented. In the first five periods, Group A implements *No Sanctions*, whereas Group B implements *Sanctions*. Starting from period 6, subjects vote every 5 periods on the rule-set to be implemented

[6]Note that in the experimental instructions (Appendix A) we refer to the *No Sanctions* (respectively, *Sanctions*) rule-sets as Rule Set 1 (respectively, Rule Set 2).

in their current group (i.e., the group in which they play in that particular period, which may differ from their initial group). Hence, every five periods, subjects first vote on the rule-set (*voting stage*), then contribute to the group account (*contribution stage*), and finally decide whether to move from their current group (*moving stage*). Information about the rule-set chosen by the majority of voters is publicly released to all subjects; the rule-set is implemented immediately after voting and applies until the next voting stage.[7] Subjects vote five times in total over the course of the experiment.

4 Setup

In this section, we present the model setup and derive the theoretical predictions arising from our setting. We consider two groups composed of n^g individuals, with $g = A, B$, denoting the group index, and where $N = n^A + n^B$ represents the total number of individuals in our economy. In each period $t \in [1, T]$, individuals play a PGG. Individual i's contribution is denoted by $C_i \in [0, E]$, where E is i's endowment, for $i = 1, ..., N$. In the *sanctioning institutions* regime, a fine equal to a fraction $s \in (0, 1]$ of the endowment not allocated to the public good applies. Each individual playing under sanctioning institutions must contribute a fixed amount c to its provision (e.g., to fund law courts and police forces). In the *Voting* treatment, individuals vote every k periods on whether to implement formal sanctions in their group (g).

Let h define individuals' "initial group"; that is, the group i is assigned to in $t = 1$, for $h = A, B$. We assume each group has $n_h = n^*$ "initial group members", with $n_A + n_B = N$.[8] We denote by $\tilde{n}^g \in [0, n^*]$ the number of members in g who are also initial group members, for $g = A, B$.

In the absence of sanctions, the per-period monetary payoff of individual i, with initial

[7] In case the two rule-sets receive the same number of votes, the rule-set for the group is randomly chosen.

[8] Whenever h and g are employed, we adopt the convention of reporting individuals' location as a superscript and individuals' initial group as a subscript.

group h and located in group g of size n^g, is

$$\pi^g_{i,h} = (E - C^g_{i,h}) + \frac{R^g_{i,h}(n^g)}{n^g} \sum_{j=1}^{n^g} C^g_j, \tag{3}$$

for $i, j = 1, ..., n^g$, and $h, g = A, B$. In (3), $r^g_{i,h}(n^g) \equiv \frac{R^g_{i,h}(n^g)}{n^g}$ represents the $MPCR$ from the public good, where $R^g_{i,h}(\cdot)$ multiplies i's share of the public good to capture the social utility of the public good provision, for $h, g = A, B$.

In accordance with our choice of parameters (see Table 1), we assume the following:

A1: $R^g_{i,h}(n^g)$ is non-decreasing in n^g and attains its maximum at $n^g = n^*$, for $h, g = A, B$.

A2: $1/n^g \leq r^g_{i,h}(n^g) \leq 1$ for $i = 1, ..., n^g$, and $h, g = A, B$.

A3: $R^h_{i,h}(n) > R^g_{i,h}(n)$ (and therefore $r^h_{i,h}(n) > r^g_{i,h}(n)$), where $R^h_{i,h}(1) < R^g_{i,h}(N)$, for $h, g = A, B$ and $h \neq g$, and for $n = 2, ..., N$.

From **A1**, the social benefit from the provision of the public good – relative to its cost – increases with the group size up to n^* and remains constant for $n^g \in (n^*, N]$. **A1** may, for instance, reflect economies of scale in the production of the public good (e.g., education). **A2** gives rise to a social dilemma: in (3), (self-interested) i has an incentive to free-ride and contribute $C^g_{i,h} = 0$, whereas efficiency requires $C^g_{i,h} = E$, for $h, g = A, B$. **A3** captures i's disutility from not residing in h. However, the last inequality in **A3** implies the "initial group premium" is not sufficiently high to compensate i for living in a largely depopulated h. Together, **A1-A3** imply a *socially efficient partition of individuals* between groups: given the same average contribution in the two groups, the public good is most beneficial to a society when each individual resides in her initial group; that is, $n^g = \tilde{n}^g = n^*$, for $g = A, B$.

When g adopts sanctions, the per-period monetary payoff of individual i, with initial group h and located in g, is given by

$$\pi^{g}_{i,h}(s) = (E - C^{g}_{i,h})(1 - s) + \frac{R^{g}_{i,h}(n^{g})}{n^{g}} \left(\sum_{j=1}^{n^{g}} C^{g}_{j} \right) - c. \tag{4}$$

We assume sanctions are *deterrent*; that is, they make it individually rational to contribute the entire endowment to the public good:

A4: $s > (1 - r^{g}_{i,h}(n^{g}))$, $\forall n^{g}$, for $h, g = A, B$.[9]

Following Markussen et al. (2014), we define:

$$\mathcal{P}^{g}_{i,h}(n^{g}) = R^{g}_{i,h}(n^{g})E - E,$$

as the "cooperation premium", that is, the increment in the earnings of i with initial group h and located in g due to the presence of formal sanctions.[10] The cooperation premium is individual-specific: it increases as (i) an individual joins her initial group, and (ii) the group size rises toward the efficient size n^{*}. We assume:

A5: $c < \mathcal{P}^{g}_{i,h}(n^{g})$ if $n^{g} > \underline{n}^{g}_{h}$, for $h, g = A, B$, where $n^{*} > \underline{n}^{g}_{h} > \underline{n}^{h}_{h}$, for $g \neq h$.

From **A5**, when individuals free-ride absent formal sanctions, implementing a formal sanction scheme generates a net benefit for an individual provided the group size is larger than the threshold $\underline{n}^{g}_{h}$. Because the sanctions' cooperation premium increases when the individual resides in her initial group, we have $\underline{n}^{g}_{h} > \underline{n}^{h}_{h}$, for $g \neq h$. In particular, given the parametriza-

[9]From Table 1, **A4** is marginally violated for individuals not located in their initial group when $n^{g} = N$, for $g = A, B$ – the difference between the left-hand side and the right-hand side in **A4** being 0.01. This discrepancy affects only one equilibrium prediction (see point (a) in the *Strong Nash Stable Equilibrium* described on page 17) in terms of the contribution of individuals not located in their initial group (because it predicts a zero contribution), but not in terms of location and voting behavior. However, we find no evidence of sanctions being non-deterrent in the case in which subjects are all located in one group. For the sake of simplicity, we therefore present a theoretical analysis based on **A4**.

[10]We define earnings as the individual monetary payoff net of the cost of formal sanctions. Notice our definition of $\mathcal{P}$ assumes individuals are self-interested and therefore do not contribute to the public good absent formal sanctions.

tion of subjects' endowments, cost of sanctioning institutions, and factors in Table 1, in our experiment we have $n^* = 5$, $\underline{n}_h^h = 2$, and $\underline{n}_h^g = 3$, for $g \neq h$.

4.1 Predictions under Selfish Preferences

We now derive the equilibrium predictions arising from our setting, assuming it is common knowledge that all individuals are self-interested and rational. We first discuss individual contribution, migration and, for VT, voting behavior. We then present equilibrium predictions for the *No-Voting* and *Voting* treatments.

4.1.1 Contribution

From **A2**, absent sanctioning institutions, individual i contributes $C_{i,h}^g = 0$, for $i = 1, ..., n^g$, and $h, g = A, B$. Also, from **A4**, individual i contributes $C_{i,h}^g = E$ in the presence of sanctions, for $i = 1, ..., n^g$, and $h, g = A, B$.

4.1.2 Migration

Let n_t^g denote the size of g in period t. Also, let $m_{i,h,t}^g$ be an indicator taking the value 1 if individual i, whose initial group is h, decides to move from her current group g in period t, and 0 otherwise, for $h, g = A, B$ and $i = 1, ..., n_t^g$.

Remark 1. Absent sanctioning institutions in both groups, an individual is indifferent between her "initial group" and the "non-initial group".

Remark 2. Suppose only group B is endowed with formal sanctions in $t + 1$. If i, for $i = 1, ..., n_t^A$, is located in her initial group A and believes $n_{t+1}^B > \underline{n}_A^B$ (resp., $n_{t+1}^B \leq \underline{n}_A^B$), then $m_{i,A,t}^A = 1$ (resp., $m_{i,A,t}^A = 0$). Similarly, if i, for $i = 1, ..., n_t^B$, is located in her initial group B and believes $n_{t+1}^B \leq \underline{n}_B^B$ (resp., $n_{t+1}^B > \underline{n}_B^B$), then $m_{i,B,t}^B = 1$ (resp., $m_{i,B,t}^B = 0$).

Remark 3. If both groups implement sanctioning institutions, every individual prefers to be located in her initial group if she believes that the other individuals will do so as well. In this case, individuals obtain the highest possible payoff.

By migrating, individuals express preferences for the public good provision and/or groups. "Voting with feet" is, however, powerless absent sanctions, because no member of either of the two groups contributes to the public good (Remark 1).

When only one group implements sanctioning institutions, different beliefs support different migration patterns. In particular, from Remark 2, because the sanctions' cooperation premium ($\mathcal{P}^{g}_{i,h}$) is a function of the group size, subjects find it individually rational to move to (or remain in) the group in which they believe *all* the other subjects will be located, regardless of the sanctioning regime implemented.[11] This statement also holds when both groups implement formal sanctions. In this case, however, each subject can find it individually rational to settle in her initial group if other subjects do so as well (Remark 3).

4.1.3 Voting

From **A5**, a threshold $\underline{n}^{g}_{h}$ exists for $h, g = A, B$, such that voting for sanctions is a dominant strategy if $n^{g} > \underline{n}^{g}_{h}$, where $\underline{n}^{g}_{h} > \underline{n}^{h}_{h}$, for $g \neq h$.

Remark 4. All else equal, when playing in her initial group, an individual enjoys a higher cooperation premium from sanctioning institutions. Therefore, individuals are more likely to vote in favor of sanctioning institutions when located in their initial group.

4.1.4 Equilibrium Outcomes

Following Robbett (2014), we focus on *Nash Stable* and *Strong Nash Stable* partitions of individuals.[12] A partition is *Nash Stable* if no individual can obtain a higher payoff from

[11] Recall that, given our choice of parameters, we have $\underline{n}^{h}_{h} = 2$ and $\underline{n}^{g}_{h} = 3$, for $g \neq h$. Hence, if the group with sanctions is (is not, respectively) individual i's initial group, i finds it optimal to be located in the group with sanctions if it contains at least two (three, respectively) members (including i). As a consequence, no individual has an incentive to be alone in a group. More specifically, each individual has an incentive to be located in the group where she believes *all* the other individuals will be located, irrespectively of whether (i) the group is her initial group, and (ii) sanctions are in place.

[12] On equilibrium selection in Tiebout models, see Greenberg and Weber (1986) and Conley and Konishi (2002).

unilaterally moving to a different group. A partition is *Strong Nash Stable* if no coalition of agents exists in which all weakly prefer *coalitionally* moving to a different group, with at least one member of the coalition being strictly better off.

***No-Voting* treatment.** When voting on institutions is not allowed and only Group B is endowed with sanctioning institutions, the following equilibria exist.

Nash Equilibria with Exogenous Institutions. *In the No-Voting treatment, two equilibria exist in which all individuals are located in either Group A or Group B. Individuals contribute to the public good only when located in the group implementing sanctions (Group B). Hence, the equilibrium in which all individuals are located in Group B is Strong Nash Stable.*[13]

Because we expect the *Strong Nash Stable* equilibrium to arise, we have the following hypothesis:

Hypothesis 1. *In the No-Voting treatment, we expect:*

(***1.1***) *Group B to become overpopulated relative to Group A by attracting A's initial members;*

(***1.2***) *subjects to contribute more when sanctioning institutions are in place;*

(***1.3***) *payoff inequality to emerge both* (*i*) *between groups and* (*ii*) *within the overpopulated group (Group B), between its initial and non-initial members.*

***Voting* treatment.** When subjects have the possibility of voting on institutions, by design only Group B implements sanctioning institutions in $t = 1, ..., k_1 - 1$, where $k_1 < T$ denotes

[13]Notice that, from the inequality $n^* > \underline{n}_h^g > \underline{n}_h^h$ in **A5**, for $h, g = A, B$ and $g \neq h$, the size of at least one group must exceed the threshold $\underline{n}_h^g$, for $h, g = A, B$. As a consequence, a (pure strategy) equilibrium in which both groups are populated does not exist. In particular, given our choice of parameters – according to which $n^* = 5$, $\underline{n}_h^h = 2$, and $\underline{n}_h^g = 3$, for $g \neq h$ – a partition of individuals such that all subjects are located in their initial group is not an equilibrium of the *No-Voting* treatment.

the first period in which there is a voting stage. Voting occurs at regular intervals of k periods: the first voting stage takes place at the beginning of period $k_1 \equiv k + 1$.

Nash Stable Equilibria. *In the Voting treatment, there exist two Nash Stable equilibria in which all individuals are located in either Group A or Group B in $t = 2$, no migration occurs in the remaining periods, all individuals vote in favor of formal sanctions in every period that includes a voting stage, and individuals contribute to the public good only when located in a group implementing sanctions.*

Initial migration driven by beliefs and, possibly, different institutional settings between groups may lead to inefficient outcomes. When all subjects are located in one group, none of them finds it individually rational to migrate further (Remark 2). The two groups then diverge in terms of welfare, as one group becomes fully depopulated. Also, payoff inequality within the populated group arises as the initial group members obtain higher returns from the public good.

Strong Nash Stable Equilibrium. *In the Voting treatment, an additional Nash Stable equilibrium exists in which:*

(a) individuals contribute 0 *(resp., E) when located in a sanction-free group (resp., in a group implementing sanctions);*

(b) all individuals are located in Group B in $t = 2, ..., k_1 - 1$;

(c) all individuals whose initial group is A move to Group A at the end of period $t = k_1 - 1$;

(d) in $t = k_1$, all individuals vote in favor of formal sanctions;

(e) no migration occurs from $t = k_1$ onward, and all individuals vote in favor of sanctioning institutions in all subsequent periods that include a voting stage.

This equilibrium is Strong Nash Stable.

This *Strong Nash Stable Equilibrium* describes the Pareto-dominant outcome that is made possible by endogenous institutional choice. Subjects initially move to the group implementing sanctioning institutions to enjoy the efficient provision of the public good. Right before the first period that includes a voting stage, individuals located outside their initial group coalitionally move and implement sanctioning institutions in their initial group. Once both groups are endowed with sanctioning institutions, this efficient partition of individuals becomes the unique *Strong Nash Stable* partition. This equilibrium also eradicates payoff inequalities both between and within groups.[14]

Because the *Strong Nash Stable Equilibrium* is Pareto-dominant, we expect this equilibrium to arise. Therefore, we have the following hypothesis:

Hypothesis 2. *In the Voting treatment, we expect:*

(***2.1***) *groups to converge toward the efficient partition of individuals; that is, the majority of subjects to be located in their initial group;*

(***2.2***) *the majority of subjects to vote for sanctioning institutions;*

(***2.3***) *subjects to contribute more when sanctioning institutions are in place;*

(***2.4***) *payoff inequality to be relatively low between groups;*

(***2.5***) *subjects located outside their initial group to be more likely to move back right before one of the periods that include a voting stage, vis-à-vis periods that do not include a voting stage.*

4.2 Predictions under Social Preferences

We briefly present some of the consequences of assigning social preferences – e.g., preferences for reciprocity and/or efficiency (Fehr and Schmidt, 1999; Charness and Rabin, 2002) –

[14] Because, in equilibrium, groups are populated by initial members only, there is no inequality within groups.

to individuals for equilibrium dynamics, as a formal analysis proves particularly cumbersome in the presence of migration.[15]

In the *Voting* treatment, social preferences can give rise to the following equilibrium predictions.

Equilibria under Social Preferences:

1. *Suppose social preferences stabilize cooperation at a sufficiently high level in both groups absent sanctioning institutions. Then, in the Voting treatment, an equilibrium exists in which individuals are efficiently located in their sanction-free initial groups.*

2. *Suppose that, absent sanctioning institutions, social preferences stabilize cooperation at a sufficiently high level in one group only. Then, in the Voting treatment, if the cost for individuals of being located away from their initial group is sufficiently high, there exists an equilibrium in which only one group implements sanctions.*

This first outcome is particularly efficient: individuals contribute to the public good in their initial group without incurring the cost of sanctioning institutions. This equilibrium can arise in the presence of a sufficiently strong aversion to advantageous inequality (Fehr and Schmidt, 1999) and/or a sufficiently strong preference for social welfare and overall efficiency (Charness and Rabin, 2002). However, given our choice of parameters and the estimates for individuals' aversion to advantageous and disadvantageous inequality reported in Fehr and Schmidt (1999), the likelihood of observing this equilibrium in our setting would be quite low.[16]

[15]For a formal discussion of the impact of social preferences on individuals' voting behavior on formal sanctioning institutions in the absence of migration, see Markussen et al. (2014).

[16]Our framework differs somewhat from the two settings – a one-period PGG with no sanctions, and a two-period PGG with peer-punishment – analyzed in Fehr and Schmidt (1999). Nevertheless, given our choice of parameters, we can obtain an estimate of the likelihood of observing an equilibrium with no sanctions by looking at a one-shot PGG. Consider then a one-shot PGG setting with no punishment and utility functions as in Fehr and Schmidt (1999). Individuals' preferences are common knowledge. Given group size $n = 5$ and $MPCR = 0.45$ (as in Table 1), the existence of a unique equilibrium with no contributions to the public good

Similarly, in the presence of a cost for individuals of being located outside their initial group, social preferences can support an equilibrium in which only one group implements sanctions. For instance, an equilibrium may exist in which each individual is located in her initial group, and only one group is able to support cooperation without implementing sanctions. Importantly, because of the cost of being located outside their initial group, individuals playing under sanctioning institutions do not find it profitable to relocate to the sanction-free group. This outcome is also relevant to the *No-Voting* treatment.

5 Results

We begin by comparing the dynamic and convergence results of the game attained when subjects can only move between groups (*No-Voting* treatment) to those attained when subjects can decide on both the rule-sets and their location (*Voting* treatment). We then examine the main determinants of subjects' contribution, migration, and voting decisions.

5.1 Dynamics and Convergence

Dynamics. We start with an overview of groups' dynamics in the two treatments. The top panels in Figure 1 plot the average size of Groups A and B over time in the *No-Voting* treatment (*NVT*). The bottom panels replicate this information for the *Voting* treatment (*VT*). The figure further discriminates between subjects whose current group is/is not their initial group.

In the *No-Voting* treatment, Group A – in which no sanctioning institutions are in place – becomes depopulated over time, whereas Group B – in which sanctions are in place – becomes overpopulated. The size of Group A goes from five members in the initial period of the experiment (on average across sessions) to about two members in the last five periods. At the same time, Group B increases from five members in the initial period to almost

is ruled out if and only if all randomly selected group members have an aversion to advantageous inequality parameter greater than 0.55. This event occurs with a probability of about 1%.

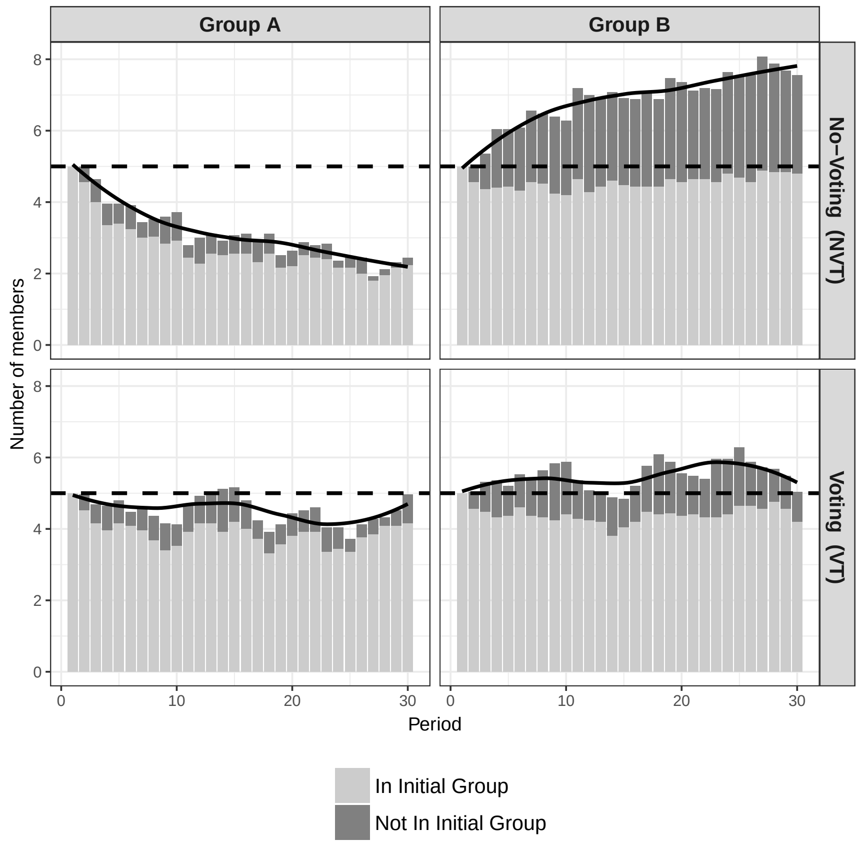

Figure 1: Distribution of participants across groups over time, by treatment. Lighter (darker) portions of the vertical bars represent the number of individuals - averaged by period across sessions - for whom their current group is (is not) their initial group. Solid lines represent locally weighted scatterplot smoothing curves fitted to the data. The dashed horizontal lines represent the situation in which participants are equally distributed between Group A and Group B.

eight members on average in the last five periods.[17] This result is broadly aligned with the findings in Gürerk et al. (2006), although the percentage of subjects eventually located in the group with sanctions is somewhat larger in their experiment (ca. 93%, against 80% in our setting). However, unlike Gürerk et al. (2006), we find little evidence that subjects who opt for sanctioning institutions early on tend to be "higher contributors": the average contribution of subjects who moved from Group A to Group B in the first period (24.44 tokens) is not significantly different from the average contribution of those who remained

[17]We computed the average size of Group A and Group B over the first and last five periods of each session, and then compared the distribution of these averages across sessions. Differences are statistically significant (two-tailed Wilcoxon signed-rank test: z=4.20, p=0.00).

in Group A (20.24) in that period (Wilcoxon signed-rank test: z=1.12, p=0.30, two-tailed). Altogether, the average size of Group A over the 30 periods of our experiment is 3.15 across all sessions, significantly lower than the average size of Group B (6.85).[18]

The upper panels in Figure 1 also reveal the composition of both groups in NVT varies throughout the experiment. In particular, Group B increases in size by attracting individuals who were initially assigned to Group A. In the first five periods, only 18.7% of the subjects initially assigned to Group A (0.94 subjects on average) are located in Group B. By the last five periods, that percentage grows to 59.4% (2.97 subjects), a statistically significant increase (Wilcoxon signed-rank test: z=4.21, p=0.00, two-tailed).[19] These results support Hypothesis 1.1.

The bottom panels in the figure show a more equal distribution of subjects between Group A and Group B in VT than in NVT. The two treatments differ both in the size and composition of the groups. First, in VT, although the average size of Group A (4.51) is significantly smaller than the average size of Group B (5.49) (Wilcoxon signed-rank test: z=-1.98, p<0.05, two-tailed), the difference between the two groups is significantly lower than in NVT (Mann-Whitney test: -4.03, p=0.00, two-tailed).[20] Furthermore, the size of the groups in VT remains quite stable over the course of the experiment: the average number of members of each group in the first five periods is not significantly different from that in the last five periods (Wilcoxon signed-rank test: z=1.20, p=0.24 for Group A, and z=-1.16, p=0.25 for Group B, two-tailed). At the same time, the average size of Group A in these last five periods (4.44) is significantly lower than 5 (Mann-Whitney test: z=-2.19, p=0.01, one-sided), suggesting that Group A finds it difficult to fully recover from the absence of sanctions in the initial periods.[21] Second, in VT, most of the members of Group A and Group B play

[18]Wilcoxon signed-rank test: z=-4.37, p=0.00, two-tailed.

[19]To assess the significance of these changes in group composition, we computed the average number of subjects initially assigned to Group A that were located in Group B over the first and last five periods of each session, and compared the distribution of these averages across sessions.

[20]The average size of the two groups in VT is statistically indistinguishable in the first five periods of the experiment, when subjects cannot vote on institutions yet (Wilcoxon signed-rank test: z=-1.01, p=0.32, two-tailed).

[21]It is also worth mentioning that, while the average size of Group B is statistically indistinguishable

in their initial groups. On average across sessions, 78% (88%) of subjects whose initial group is Group A (B) play in Group A (B), and these proportions do not vary significantly over the 30 periods of the experiment: the p-value for a permutation test for repeated measures (Pesarin and Salmaso, 2010) is 0.11. Altogether, 83% of subjects are located in their initial group in any given period, in accordance with Hypothesis 2.1. The corresponding percentage for NVT (72.7%) is significantly lower.[22]

To better understand the differences in the size and composition of groups between NVT and VT, Figure 2 looks at subjects' migration behavior. The left panel plots the average share of subjects moving between groups in each period for the two treatments. The right panel distinguishes between moves toward and moves away from groups that are implementing sanctions at the time these moves take place.[23]

As shown in the left panel of the figure, the percentage of subjects who move between groups in each period is systematically higher in NVT than in VT. As a summary measure of the differences in the propensity to migrate between treatments, we compute the average proportion of subjects who switch groups in NVT and VT over the course of the experiment (i.e., across the 30 periods). In NVT, 14.76% of individuals migrate in any given period, on average, while 11.52% move between groups in VT. This difference is statistically significant (Mann-Whitney test: z=1.97, $p<0.05$, two-tailed).[24] The right panel, in turn, shows subjects are more likely to migrate toward groups that have sanctioning institutions in place at the time of the move. Considering only those periods/sessions in VT in which the two groups

between the two treatments in the first five periods of the experiment (Mann-Whitney test: z=0.80, p=0.43, two-tailed), the mean number of Group B members in VT becomes significantly larger than in NVT already in period 6, right after the first voting stage (z=1.99, $p<0.05$, two-tailed).

[22] We computed the average number of subjects located in their initial group in each VT and NVT session, and compared their distribution across the 25 sessions ran for each treatment. Mann-Whitney test: z=-3.23, p=0.00, two-tailed.

[23] Recall that, in periods that include a voting stage in VT, subjects vote on the rule-set at the beginning of the period, and decide whether to migrate or not at the end of the same period. In the right panel of Figure 2, groups have already adopted sanctioning institutions prior to the moves, and these sanctions are still in place at the moving stage - i.e., when subjects decide whether to switch groups or not.

[24] We also compared the distribution of the average proportion of movers under NVT and VT across sessions using a permutation test, which imposes fewer assumptions and has higher power than the Mann-Whitney test (Pesarin and Salmaso, 2010). The null hypothesis of no difference between treatments is again rejected (p-value=0.01, two-tailed).

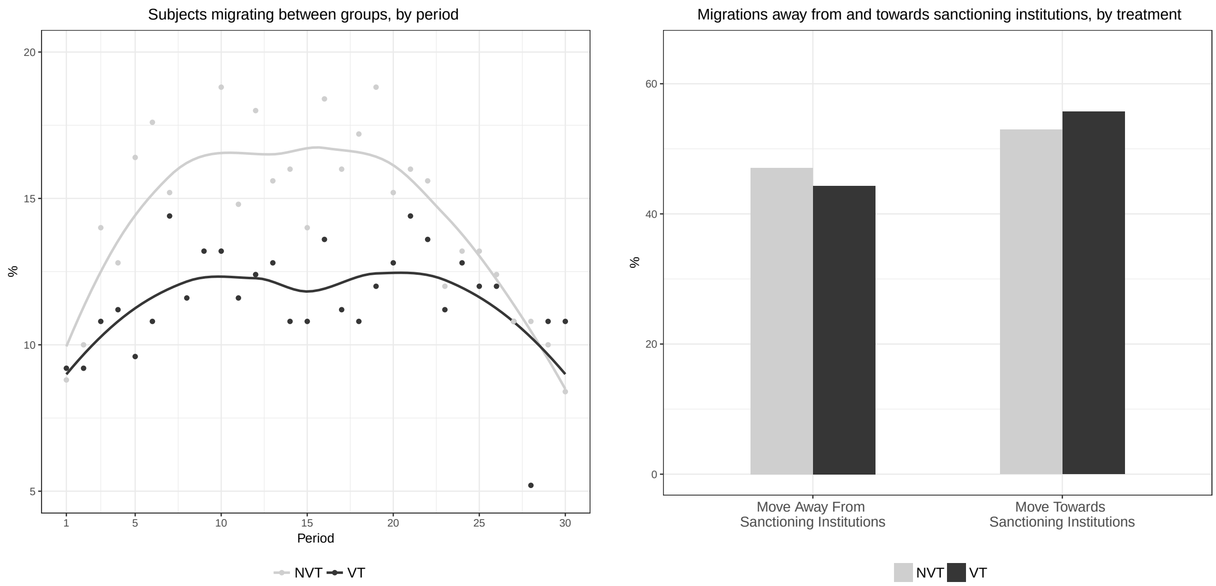

Figure 2: Migration patterns under the *Voting* and *No-Voting* treatments. The left panel plots the percentage of moves between groups over time. Black (gray) dots represent the proportion of moves in each period - averaged across sessions - for VT (NVT); solid lines represent locally weighted scatterplot smoothing curves fitted to the data. The right panel plots the proportion of moves away from/toward groups that are implementing sanctions in the period in which the moves take place – out of the total number of group shifts that occur in that period across all sessions – under VT (black bars) and NVT (gray bars). For VT, the sample in the right panel comprises only periods-sessions in which the rule-sets implemented by Group A and Group B differ at the time of the moves.

are implementing different rule-sets (ca. 41% of the observations), we find 55.7% of the movements go from groups that are sanction-free at the time of the switch towards groups that are implementing sanctions at that same time, while 44.3% of the movements go in the opposite direction. In NVT, the corresponding figures are 52.9% and 47.1%, respectively. In both treatments, the proportion of movements toward sanctioning institutions is significantly higher than the proportion of movements away from institutions (exact binomial test for the distribution of average proportions in each session: p=0.02 for VT, p=0.05 for NVT, two-tailed).[25]

[25] Also, considering only periods/sessions in which the two groups have different institutions, the proportion of switches towards groups that are implementing sanctions at the time of the move in VT is not significantly different from the corresponding proportion in NVT (χ^2=0.79, p=0.37 for a two-sided test for equality of proportions).

Overall, whereas the right panel of Figure 2 points to quite similar migration choices in *NVT* and *VT*, Figure 1 shows the dynamics in terms of the size and composition of the groups differ substantially between treatments. We attribute this difference to the ability of subjects to vote for their preferred institutions in *VT*.

Next, we analyze individuals' contribution to the public good over time, because this analysis can help account for subjects' propensity to move toward groups endowed with sanctions. As Figure 3 shows, in both *NVT* and *VT*, the average contribution in sanction-free groups is significantly lower than in groups implementing sanctions. In *NVT*, average contributions across sessions are 45.67 tokens for groups with institutions and 20.55 tokens for sanction-free groups; in *VT*, the corresponding values are 46.61 and 24.75. Differences in the average contribution per session between sanctioning and sanction-free groups are statistically significant (Wilcoxon signed-rank test: z=4.37, p=0.00 for *NVT*, and z=4.37, p=0.00 for *VT*, two-tailed), consistent with Hypotheses 1.2 and 2.3.

The average contributions of groups implementing sanctions are not significantly different across treatments (Mann-Whitney test: z=0.78, p=0.44, two-tailed). By contrast, the average contribution of sanction-free groups is somewhat larger in *VT* than in *NVT* (Mann-Whitney test: z=1.93, p=0.05, two-tailed).[26,27] Leaving aside the potential self-selection of subjects into groups with different institutions – which can occur in both treatments – this latter result may be due to two inter-related reasons. First, in principle only groups with sufficiently high contributions can survive without implementing sanctions in *VT*. Second, in line with Sutter et al. (2010), subjects tend to contribute more when they choose not to have sanctions, as opposed to the case in which the same institutional setting is exogenously imposed.[28]

[26] This result bears some similarities with Ertan et al. (2009) and Sutter et al. (2010).

[27] In both treatments, the average contribution in sanction-free groups is significantly greater than zero (z=6.48, p=0.00 in *NVT*; z=6.48, p=0.00 in *VT*, one-tailed Mann-Whitney tests).

[28] Our data seems to provide support for both explanations. In *VT*, the average contribution in period 5 among those sanction-free groups that go on to vote against institutions in the first voting stage (at the beginning of period 6) is 31.71 tokens – significantly higher than the average period-5 contribution of sanction-free groups in *NVT* (22.19) (Mann-Whitney test: z=1.98, p<0.05, two-tailed). This is consistent with our first conjecture. Additionally, among groups in *VT* that vote against sanctions in any given voting stage, the

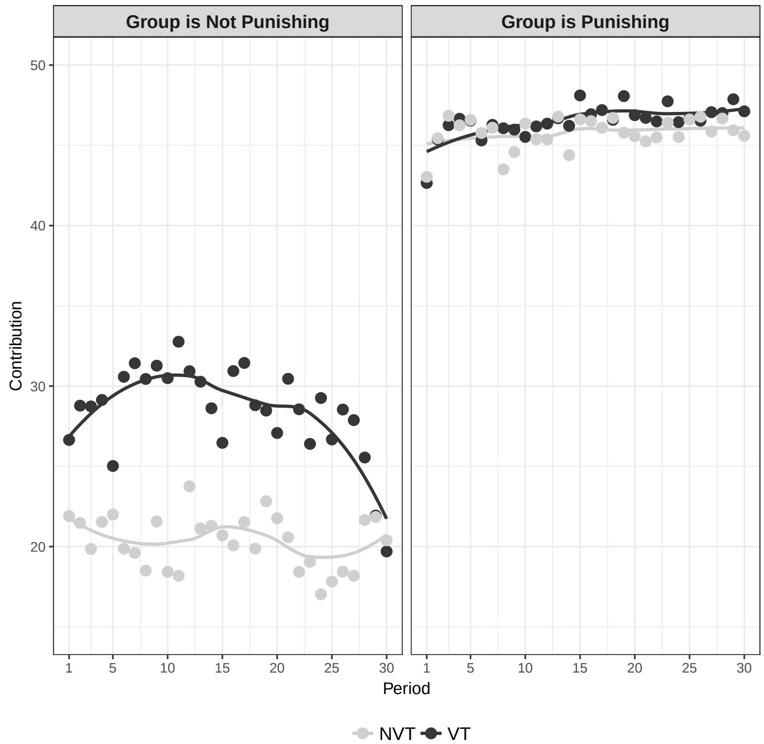

Figure 3: Average group contribution per period, by institutional setting and treatment. The left (right) panel plots the average contribution for groups without (with) sanctioning institutions. In both cases, black (gray) dots represent the mean contribution per period - averaged across sessions - under VT (NVT); solid lines represent locally weighted scatterplot smoothing curves fitted to the data.

Finally, Figure 4 analyzes subjects' average payoffs over time in the two treatments.[29] In NVT, the mean per-period payoffs (averaged across sessions) are 67.30 for Group A members and 82.17 for Group B members, a statistically significant difference (Wilcoxon signed-rank test: z=4.34, p=0.00, two-tailed).[30] Furthermore, among subjects located in Group B, the

average contribution in the period right after the voting stage is 22.07 tokens. This is again significantly higher than the average contribution of sanction-free groups in NVT for the same periods (19.10) (Mann-Whitney test: z= 2.91, p= 0.00, two-tailed), which lends credence to our second conjecture.

[29]More specifically, for every period, Figure 4 plots the average payoff for subjects located in Groups A and B. To complement the information reported in Figures 3 and 4, Figure A.1 in the Appendix analyzes the average payoffs per period in groups with and without sanctions. The conclusions emerging from this comparison are analogous to those drawn from the analysis of average contributions in sanction-free and sanctioning groups (Figure 3).

[30]These average payoffs are below those predicted by the *Strong Nash Stable* equilibrium under selfish preferences (77.3 for Group A and 97.5 for Group B) described in Section 4.1. Social preferences might help account for the difference between observed and predicted payoffs, because subjects may attempt to cooperate absent sanctions rather than (immediately) move to the sanctioning group.

average payoff of initial Group B members (87.84) is significantly higher than that of non-initial members (72.41) (Wilcoxon signed-rank test: z=4.37, p=0.00, two-tailed).[31] Both results support Hypothesis 1.3.

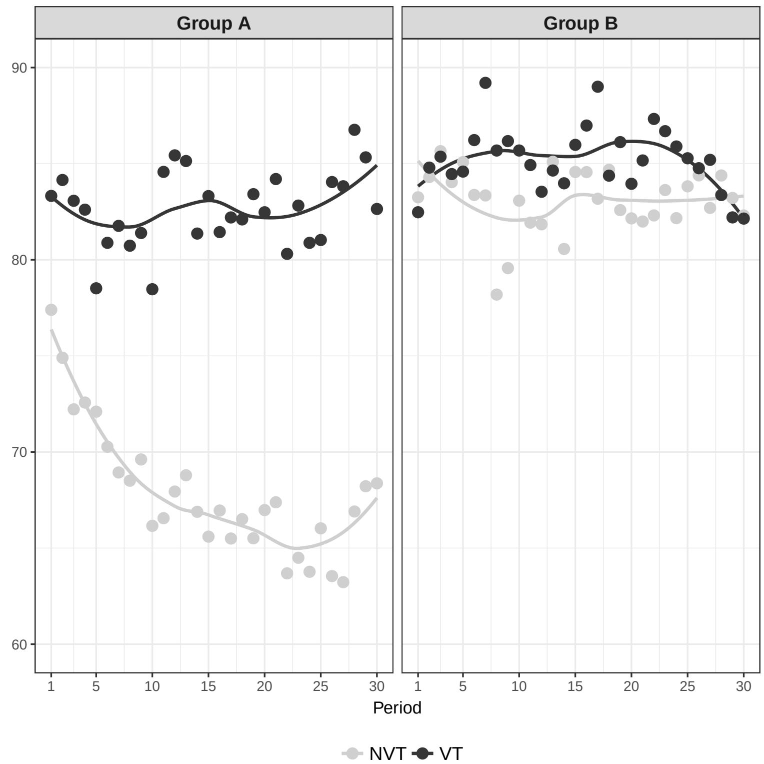

Figure 4: Average participant payoff per period, by current group and treatment. Black (gray) dots represent the mean payoff per period (averaged across sessions) of subjects located in Groups A and B under *VT* (*NVT*). Solid lines represent locally weighted scatterplot smoothing curves fitted to the data.

In *VT*, the average per-period payoffs across sessions are 81.18 tokens for Group A members and 85.15 tokens for Group B members.[32] This difference is statistically significant (Wilcoxon signed-rank test: z=2.27, p=0.02, two-tailed), somewhat contradicting Hypothesis 2.4.[33] Nevertheless, the disparity in payoffs between groups in *VT* is significantly lower

[31] Values of observations for this comparison were computed as follows: for each subject ever located in Group B, we added her payoffs during all the periods she spent in Group B - and only during those periods - and divided them over the total number of periods the subject spent in Group B over the course of the experiment. We thus obtained a single average per-period payoff for the individual. We then distinguished between the subset of these individuals who were initially assigned to Group B and those who were not.

[32] These payoffs are again smaller than those predicted by the *Strong Nash Stable* equilibrium under selfish preferences (91.92 for Group A and 97.5 for Group B).

[33] This comparison considers the average payoffs for members of Group A and B over the 30 periods of

than in *NVT* (Mann-Whitney test: z=-4.26, p=0.00, two-tailed), suggesting that subjects' ability to vote on sanctioning institutions leads to more equal societies. Moreover, the average payoff across both groups in *VT* (84.07) is significantly higher than in *NVT* (78.47) (Mann-Whitney test: z=2.80, p=0.00, two-tailed). Although the average payoffs across groups are below the predictions of the *Strong Nash Stable* equilibria (equal to 94.71 for *VT* and 87.42 for *NVT*), our results are in line with the theoretical prediction according to which subjects' ability to vote for institutions increases efficiency.

In the following subsection we analyze the stability of our results by investigating groups' convergence.

Convergence. We focus on two main dimensions of convergence. We define "full convergence in groups by period γ" as the situation in which, given a particular composition of groups in a given session, no subject moves between groups from γ onward. We also adopt a less restrictive definition, "near convergence in groups", which allows for at most one subject switching groups each period from γ onward. For the second dimension, we define "convergence in institutions by period τ" as the situation in which neither group changes its institutional setting from τ onward in a given session.

The upper and middle parts of Table 2 show, for each treatment, the percentage of sessions achieving full and near convergence in groups, respectively, at least three periods before the end of the experiment (i.e., by period $\gamma = 27$ or earlier). For each fully/nearly converging session we also compute the convergence time, defined as the period number after which the corresponding convergence criterion continues to hold. Table 2 reports the average convergence time across all sessions that achieve full/near convergence in groups.[34] We also report the average percentage of subjects who play in their initial group from the

the experiment. The difference between groups remains significant when we remove the first five periods in which subjects cannot decide on the rule-sets (Wilcoxon signed-rank test: z=2.38, p=0.02, two-tailed). By contrast, average payoffs are statistically indistinguishable between groups in those initial periods before the first voting stage (Wilcoxon signed-rank test: z=1.60, p=0.11, two-tailed).

[34]Note that the average convergence time is lower than 27, since most of the sessions that achieve full/near convergence in groups *by* γ do so before this period.

convergence period onward.

Table 2: Convergence in groups and convergence in institutions

	No-Voting	*Voting*
Full Convergence in Groups ($\gamma = 27$)		
Converging Sessions (%)	20.00	32.00
Average Convergence Time	25.00	22.50
Subjects in Initial Group (%)	68.00	100.00
Near Convergence in Groups ($\gamma = 27$)		
Converging Sessions (%)	52.00	48.00
Average Convergence Time	22.92	19.75
Subjects in Initial Group (%)	68.97	96.39
Convergence in Institutions ($\tau = 22$)		
Converging Sessions (%)		68.00
Average Convergence Time		11.00
"Two Groups with Sanctions" (%)		52.00
"Two Groups without Sanctions" (%)		24.00
"One Group with/without Sanctions" (%)		24.00

Notes: The upper and middle parts of the table report the percentage of sessions achieving full/near convergence in groups, the average time of (full/near) convergence across all the converging sessions, and the average percentage of subjects playing in their initial group from the time of convergence onward. The lower part presents the percentage of sessions that converge in institutions, the average time of convergence across all the converging sessions, and the percentage of sessions that converge to each possible combination of institutions in the two groups. Since subjects in the *No-Voting* condition cannot choose the institutions governing their interactions, we only report results for (full and near) convergence in groups for this treatment. Convergence time is defined as the period number following which the relevant criterion of convergence - i.e., full/near convergence in groups, convergence in institutions - continues to hold in a given session. This value is then averaged across all those sessions that achieved full/near convergence in groups (upper/middle panels) and in institutions (bottom panel). For illustration purposes, we present results for $\gamma = 27$ and $\tau = 22$; the main findings remain similar for alternative values of γ and τ.

In *NVT*, 52% of the sessions achieved near convergence in groups by period 27, and 20% achieved full convergence. On average, sessions that fully converge in groups do so in 25 periods; the average time for near convergence is somewhat shorter (ca. 23 periods). In line with the predictions of the *Nash Equilibria with Exogenous Institutions* (see Section 4.1.4), the groups' composition is far from efficiency: less than 70% of subjects play in their initial group from the time of full/near group convergence onward.

In *VT*, we observe somewhat similar results to those in *NVT* in terms of the percentage of fully/nearly converging sessions (32/48%) and their average convergence time (19.75/22.5 periods). However, when full convergence occurs, the two groups are entirely populated by their initial members. Even allowing for near group convergence, a whopping 96.39% of subjects in *VT* play in their initial group. In line with the predictions of the *Strong Nash Stable Equilibrium*, this finding points to subjects' ability to converge toward the efficient partition of individuals between groups in *VT*. These results are similar for alternative values of γ.

The lower part of Table 2 displays the results for the convergence in institutions. For this type of convergence, we report the results for *VT* only, because subjects cannot vote in *NVT*. The table shows that almost 70% of the sessions converge in institutions by period $\tau = 22$ (i.e., the period in which the next-to-last voting stage takes place). The average convergence time across all those sessions that converged in institutions is 11 periods - i.e., groups in those converging sessions maintain their rule-sets unchanged from period 11 onward on average. Results are again robust to different values of τ.

The table also reports the percentage of sessions that converge to each possible combination of institutions in the two groups. In the majority of the cases (52%) in which there is convergence in institutions by period 22, both groups converge to implementing sanctions. This result is consistent with the predictions of the *Strong Nash Stable Equilibrium* under selfish preferences and, in particular, supports Hypothesis 2.2.

We also observe 24% of sessions converging to both groups not implementing sanctions. Although this outcome is in line with the first *equilibrium under social preferences* described in Section 4.2, its frequency appears surprisingly high given the low levels of cooperation observed by Fehr and Schmidt (1999) in a one-shot PGG game with no punishment (see footnote 16). However, we believe that our (repeated) framework – in which subjects can vote to implement sanctions – bears *some* similarities with the "two-stage PGG with peer-punishment" also analyzed in Fehr and Schmidt (1999). There, the authors find that cooperation is more likely to be sustained than in a one-shot PGG without sanctions. Therefore, we also expect an equilibrium in which cooperation is sustained in two sanction-free groups to occur with a non-negligible probability.

Finally, 24% of the sessions converge to two groups implementing different institutions (i.e., only one group implements sanctions). These sessions behave in accordance with the second *equilibrium under social preferences* described in Section 4.2.

5.2 Individual Behavior

To better understand the mechanisms underlying the group-level patterns reported above, this section presents individual-level analyses of contribution, migration, and voting decisions, focusing primarily on our main treatment (VT).[35]

Contribution. The first column of Table 3 reports the marginal effects from a (doubly) censored regression model (Greene, 2012) in which the dependent variable, $Contribution_{i,t}$, is the individual contribution of subject i in period t under VT.[36] We have the following explanatory variables: $Sanctions_{g,t}$, a binary covariate that equals 1 if a sanctioning mechanism is in place in i's current group g, and 0 otherwise; $Initial\ Group_{i,t}$, a dummy taking the value 1 if i is located in the group she was assigned to at the beginning of the experiment, and 0 otherwise; $Group\ Contribution_{g[-i],t-1}$, the average contribution of the other members of i's current group in period $t-1$; and *abs(Group Size*$_{g,t}$ *- 5)*, the absolute value of the difference between 5 and the number of members in i's current group, measuring the impact of deviations from the "optimal" group size on individual contributions. The model also includes subject and period random effects to account for time-invariant individual heterogeneity and common temporal shocks affecting all subjects, along with session and group

[35]For ease of exposition, we estimate the models and present results for each of these outcomes separately. The findings reported below are generally similar if we fit panel dynamic simultaneous equations models, even though standard estimation approaches (e.g., Akashi and Kunitomo, 2012; Hsiao and Zhou, 2015) require ignoring the censoring of contributions and the dichotomous nature of migration and voting decisions.

[36]"Raw" parameter estimates – representing the effect of the predictors on the uncensored latent variable – are reported in the Appendix (Table A.1).

random intercepts to control for potential session effects (Fréchette, 2012) and contemporaneous correlation between same-group members (Poen, 2009).[37]

Individual contributions to the public account increase significantly when subjects belong to a group that implements sanctioning institutions. This result suggests participants generally understood the mechanism at play and responded to incentives, and reinforces the group-level evidence in support of Hypothesis 2.3. Also, all else equal, average individual contributions are significantly higher when subjects play in their initial group. This finding is again consistent with incentives, because the returns to contributions to the public account are higher when participants remain in their initial group than when they are in the other group. Participants also tend to contribute more after observing higher contribution levels from their fellow group members in the previous period. On the other hand, i's contribution drops when the size of her group deviates from 5.

Column (2) repeats the specification in column (1), but accommodates possible asymmetric impacts that deviations from the optimal group size might have on subjects' contributions. To this end, we separate *abs(Group Size*$_{g,t}$ *- 5)* into $\mathbb{1}$(Group Size<5) (5 - Group Size$_{g,t}$) and $\mathbb{1}$(Group Size>5) (Group Size$_{g,t}$ - 5), where $\mathbb{1}(A)$ is the indicator function taking the value 1 if A is true, and 0 otherwise. The coefficients for both variables are negative and statistically significant, indicating that both negative and positive deviations from 5 are systematically associated with a drop in individual contributions.

[37] The results are generally similar for fixed-effects censored regression models estimated using Dhaene and Jochmans (2015)'s split-panel jackknife approach to correct for incidental parameter bias, as well as using Honoré (1992)'s semiparametric estimator. See Table A.2 in the Appendix.

Table 3
Determinants of individual contributions

	(1)	(2)	(3)	(4)	(5)
Sanctions$_{g,t}$	13.86***	13.82***	12.93***	10.82***	10.50***
	(0.29)	(0.29)	(0.29)	(1.30)	(1.24)
Initial Group$_{i,t}$	1.73***	1.77***	1.36***	0.51*	0.37
	(0.31)	(0.32)	(0.31)	(0.31)	(0.32)
Group Contribution$_{g[-i],t-1}$	0.13***	0.13***	0.03**	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
abs(Group Size$_{g,t}$ *- 5)*	-0.36***		-0.29***	0.39***	
	(0.10)		(0.09)	(0.10)	
$\mathbb{1}$*(Group Size*$_{g,t}$$<$*5)(5 - Group Size*$_{g,t}$*)*		-0.47***			1.01***
		(0.16)			(0.16)
$\mathbb{1}$*(Group Size*$_{g,t}$$>$*5)(Group Size*$_{g,t}$ *- 5)*		-0.33***			0.11
		(0.11)			(0.12)
Contribution$_{i,t-1}$			0.14***	0.12***	0.12***
			(0.01)	(0.01)	(0.01)
Observations	7,250	7,250	7,250	7,250	7,250
Log likelihood	-24,993.46	-24,987.21	-23,777.47	-24,246.43	-24,215.80

Notes. The table reports marginal effects for the covariates included in the panel models for individual contributions; units of observation are individuals-per-period. Columns (1)-(3) are fitted to data from the *Voting* treatment; columns (4)-(5) use data from the *No-Voting* condition. All specifications include subject, period, group, and session random effects. Standard errors are presented in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%.

The negative marginal effect of $\mathbb{1}$(Group Size<5) (5 - Group Size$_{g,t}$) is likely due to the fact that the factor multiplying total contributions to the public good becomes lower as the size of group continues to decrease below 5 (see Table 1). The negative impact of $\mathbb{1}$(Group Size>5) (Group Size$_{g,t}$ - 5) on $Contribution_{i,t}$, in turn, is consistent with the "bad apple" hypothesis (Marwell and Schmitt, 1972). Specifically, Marwell and Schmitt (1972) argue that individuals will be willing to cooperate as long as other members of the group cooperate, and that the probability of the existence of a non-cooperator ("bad apple") rises with the group size. This may thus explain why we observe a drop in individual contributions as the number of members increases (above the optimal size).

To account for autocorrelation and for subjects' tendency to adjust their contributions over time (Smith, 2013), column (3) incorporates the lag of the dependent variable among the regressors of the column (1) specification. The estimate for $Contribution_{i,t-1}$ indicates that roughly 14% of subject i's contribution in period t is explained by her contribution in the previous period. Nonetheless, the substantive conclusions emerging from previous models remain unchanged, although the association between $Contribution_{i,t}$ and the other explanatory variables – and in particular, the marginal effect of $Group\ Contribution_{g[-i],t-1}$ – becomes weaker.

Consistent with the results for VT, the estimates in column (4) – which replicates the column (3) specification for NVT – show again significant persistence in individuals' propensity to cooperate, which is further strengthened by sanctioning institutions. However, the results in column (4) differ from those of column (3) in two respects. First, when subjects are not allowed to choose the rule-set governing their interactions, previous contributions of other members of i's current group do not have a systematic influence on $Contribution_{i,t}$. A possible explanation for this discrepancy is related to the fact that subjects tend to move more in NVT than in VT (see Section 5.1). As a consequence, individuals may place less weight on previous group members' contributions in NVT because they expect their group composition to change more frequently. Second, the impact of deviations from the optimal group size in column (4) goes in the opposite direction. To shed light on this result, column (5) again divides *abs(Group Size$_{g,t}$ - 5)* into $\mathbb{1}$(Group Size<5) (5 - Group Size$_{g,t}$) and $\mathbb{1}$(Group Size>5) (Group Size$_{g,t}$ - 5). We find that only negative deviations from the optimal group size are significantly associated with a rise in contributions in NVT. This suggest that, when unable to change institutions, subjects tend to increase their contributions in order to attract other individuals to their depopulating group.

The main conclusions emerging from Table 3 also hold if data from both treatments are pooled and interactions between the explanatory variables and an indicator for VT are added to this specification (see Table A.3 in the Appendix).[38]

Migration. We now study the main determinants of subjects' migration decisions. Column (1) of Table 4 reports marginal effects obtained from a panel probit model fitted to data from our main treatment (VT). The dependent variable is $Migration_{i,t}$, a dummy taking the value 1 if subject i moved between groups in period t, and 0 otherwise. We include as predictors of the moving decision the same set of explanatory variables used in the analysis of individual contributions, replacing $Group\ Contribution_{g[-i],t-1}$ with $Contribution_{i,t}$ - $Group\ Contribution_{g[-i],t}$, the difference between i's contribution and the average contribution of the other group members in period t. Additionally, we control for the potential influence of the other group's rule-set ($Sanctions_{h\neq g,t}$) and average contribution levels ($Group\ Contribution_{h\neq g,t}$) on i's moving choices.[39] The model also incorporates subject-specific (correlated) random effects (Wooldridge, 2005, 2010) as well as period, group, and session random intercepts.[40]

[38]We also account for the potential endogeneity of covariates and allow for dynamic feedback from the outcome to the explanatory variables. Table A.4 in the Appendix replicates the analyses in Table 3 using Fernández-Val and Vella (2011)'s two-step estimator for non-linear panel data models, taking individuals' contributions in the first activity and pre-determined regressors as instruments. The main results remain qualitatively similar.

[39]As mentioned in Section 3, individuals receive information about the average contribution to the public good and the rule-set implemented in the other group prior to the moving stage.

[40]Parameter estimates are reported in the Appendix (Table A.5). For robustness, we also fitted two-way (subject and period) fixed-effects probit models using the method developed by Fernández-Val and Weidner (2016). The estimates are aligned with those from the random effects models (see Table A.6 in the Appendix).

Table 4
Determinants of individual migration decisions

	(1)	(2)	(3)	(4)	(5)
Sanctions$_{g,t}$	-8.04***	-8.07***	-8.22***	-8.27***	-21.96***
	(1.15)	(1.14)	(1.12)	(1.10)	(2.78)
Initial Group$_{i,t}$	-23.49***	-23.43***	-23.21***	-22.12***	-18.00***
	(1.84)	(1.85)	(1.90)	(1.93)	(1.59)
Contribution$_{i,t}$ - *Group Contribution*$_{g[-i],t}$	3.15***	3.14***	3.12***	3.12***	1.87***
	(0.41)	(0.42)	(0.39)	(0.40)	(0.41)
abs(Group Size$_{g,t}$ *- 5)*	1.70***	1.72***	1.73***	1.73***	1.07***
	(0.30)	(0.30)	(0.30)	(0.31)	(0.34)
Sanctions$_{h \neq g,t}$	3.77***	3.63***	3.63***	3.67***	
	(1.01)	(1.01)	(1.01)	(1.00)	
Group Contribution$_{h \neq g,t}$	0.12	0.21	0.20	0.23	0.08
	(0.42)	(0.41)	(0.41)	(0.41)	(0.49)
Vote Different from the Group$_{i,t-1}$		6.25***			
		(1.95)			
Vote Different from the Group$_{i,t-1}$ *under No Sanctions*			4.47	4.36	
			(3.32)	(3.22)	
Vote Different from the Group$_{i,t-1}$ *under Sanctions*			7.50***	7.90***	
			(2.40)	(2.44)	
Period Before a Voting Stage$_t$ - *Initial Group*$_{i,t}$				-0.68	
				(0.95)	
Period Before a Voting Stage$_t$ - *Outside Initial Group*$_{i,t}$				7.56**	
				(3.38)	
Observations	7,250	7,250	7,250	7,250	7,250
Log likelihood	-2,079.81	-2,070.96	-2,069.19	-2,066.51	-2,568.14

Notes. The table reports the change in the probability of moving (in percentage points) associated with a change in the covariates; units of observation are individuals-per-period. Columns (1)-(4) are fitted to data from the *Voting* treatment, while Column (5) uses data from the *No-Voting* condition. All the models include subject, period, group, and session random effects. Standard errors are presented in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%.

All else equal, subjects are less likely to move if they are located in their initial group and if their current group has adopted a sanctioning institution. On the other hand, the larger the difference between i's contribution and the average contribution of other members of her group, the higher the probability that she moves. The probability that the average subject moves in any given period also rises when the other group is implementing sanctions, and when the number of members of her group deviates from the optimal size.[41]

Column (2) incorporates a dummy variable, *Vote Different from the Group*$_{i,t-1}$, measuring whether i's institutional choice in the most recent period in which a voting stage took place differed from the decision of the majority of her group. We observe that the probability of migrating increases when the group did not adopt subjects' preferred rule-set in the most recent voting stage.

To determine whether this positive effect is contingent on the institutional setting implemented, column (3) adds an interaction between *Vote Different from the Group*$_{i,t-1}$ and *Sanctions*$_{g,t}$ to the previous specification. We can therefore compute the marginal effects of *Vote Different from the Group*$_{i,t-1}$ *under No Sanctions* (i.e., when *Sanctions*$_{g,t}$=0) and *Vote Different from the Group*$_{i,t-1}$ *under Sanctions* (when *Sanctions*$_{g,t}$=1). Although the probability of moving is always higher on average for "dissenters" than for subjects who agreed with the majority decision, only those who voted against sanctioning in groups that adopted centralized punishment are significantly more likely to migrate.

In column (4), we add an interaction between *Period Before a Voting Stage*$_t$ – a dummy for periods immediately preceding one of the five periods in which a voting stage takes place – and *Initial Group*$_{i,t}$. This allows us to assess whether subjects in/out of their initial group are more or less likely to move just before groups choose their institutions.[42] Specifically, we calculate the marginal effects of *Period Before a Voting Stage*$_t$ when subjects are located in their initial group (setting *Initial Group*$_{i,t}$=1 for all i) and when subjects are located outside

[41] Deviations both above and below 5 are statistically significant, although the marginal effect on migration is somewhat larger when the size of the group is below the optimal size.

[42] Recall that, in periods in which a voting stage occurs, the vote takes place before the moving stage.

their initial group (i.e., when *Initial* $Group_{i,t}$=0 for all i). In accordance with Hypothesis 2.5, we find that subjects located outside their initial group are more likely to move back right before a period that includes a voting stage – presumably to affect the outcome of the vote – than in other periods. By contrast, for individuals located in their initial group, no significant differences exist in the probability of moving right before a period that starts with a voting stage vis-à-vis other periods. Also, we find no evidence of individuals being more likely to return to their original group right after a voting stage, once the institutional choice is known.[43]

For completeness, column (5) repeats the specification in column (1) using data from the *No-Voting* treatment. The marginal effects of the covariates are qualitatively similar to those for VT, although the impact of the sanctioning institution on the probability of moving is more than one and a half times larger – in absolute value – when subjects are not allowed to vote on the rule-set.[44]

Voting. Lastly, we examine the determinants of subjects' voting decisions. Table 5 reports marginal effects computed from random effects panel probit models fitted to data from the five periods in VT that include a voting stage.[45] The dependent variable, $Vote_{i,t}$, equals 1 (0) if i voted for (against) the sanctioning institution in period t.

The explanatory variables in column (1) are essentially the same as those of previous individual-level analyses, with $Sanctions_g$ and $Sanctions_{h\neq g}$ now lagged one period to capture whether a punishment scheme is in place in any of the two groups before the current voting

[43]Additionally, Figure A.2 in the Appendix compares the probability of moving from sanction-free to sanctioning groups in periods before a voting stage vis-à-vis other periods, as a function of the average contribution in the group without sanctions. The figure shows that, also in periods other than those preceding a voting stage, subjects tend to migrate towards sanctioning institutions when contributions in the group without sanctions are low. Nonetheless, and consistent with the results in Table 4, the probability of moving is highest right before a vote, particularly among subjects located outside their initial group.

[44]$Sanctions_{h\neq g,t}$ is dropped from this specification due to its perfect collinearity with $Sanctions_{g,t}$. Table A.7 in the Appendix replicates the analyses in Table 4 using Fernández-Val and Vella (2011)'s two-step estimator to account for potential endogeneity of the explanatory variables, taking individuals' contributions in the first activity and pre-determined regressors as instruments. Although the marginal effect of *abs(Group* $Size_{g,t}$ *- 5)* is never significant, the estimates for the other covariates remain similar.

[45]Parameter estimates are reported in Table A.8 of the Appendix. For completeness, Table A.9 presents estimates from bias-corrected fixed-effects panel probit models (Fernández-Val and Weidner, 2016).

stage. We also include an interaction between $Sanctions_{h \neq g,t-1}$ and $Group\ Contribution_{h \neq g,t-1}$ to assess whether the average contribution in the other group affects i's voting decisions and if such influence is contingent on the rule-set implemented by this other group.[46]

Table 5
Determinants of individual voting decisions.

	(1)	(2)	(3)	(4)
$Sanctions_{g,t-1}$	16.11***	15.94***	18.01***	14.89***
	(4.06)	(4.20)	(4.97)	(3.36)
$Initial\ Group_{i,t}$	11.98***	12.46***	14.31***	8.25**
	(3.88)	(3.97)	(4.94)	(3.66)
$Contribution_{i,t-1}$ - $Group\ Contribution_{g[-i],t-1}$	-1.81	-1.85	-2.33	0.34
	(1.49)	(1.54)	(1.92)	(1.32)
$abs(Group\ Size_{g,t}$ - $5)$	2.52*			
	(1.37)			
$Sanctions_{h \neq g,t-1}$	16.50***	14.84***	17.92***	15.40***
	(6.11)	(5.10)	(7.04)	(4.70)
$Group\ Contribution_{h \neq g,t-1}$ if $Sanctions_{h \neq g,t-1} = 0$	-6.29***	-5.87***	-6.58***	-6.54***
	(2.06)	(2.02)	(2.45)	(1.90)
$Group\ Contribution_{h \neq g,t-1}$ if $Sanctions_{h \neq g,t-1} = 1$	5.81	5.45	4.34	1.72
	(4.82)	(4.80)	(5.65)	(3.66)
$\mathbb{1}(Group\ Size_{g,t} < 5)(5$ - $Group\ Size_{g,t})$		3.03**	3.66**	1.06
		(1.44)	(1.70)	(1.26)
$\mathbb{1}(Group\ Size_{g,t} > 5)(Group\ Size_{g,t}$ - $5)$		1.42	1.30	0.88
		(2.04)	(2.51)	(1.88)
$Time\ Trend$			4.52***	3.02***
			(1.05)	(1.16)
$Vote_{i,t-1}$				45.87***
				(3.21)
Observations	1,250	1,250	1,250	1,000
Log likelihood	-648.06	-647.78	-641.67	-488.31

Notes. The table reports the change in the probability of voting for sanctions (in percentage points) associated with a change in the covariates; units of observation are individuals-per-period (with the sample restricted to the five periods that include a voting stage). All the columns report estimates obtained under the *Voting* treatment (subjects cannot vote in NVT). All specifications include subject, period, group, and session random effects; columns (3)-(4) also include a linear time trend. Standard errors are presented in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%.

Consistent with Remark 4, the probability of voting in favor of the sanctioning institution is systematically higher for subjects who remain in the group to which they were originally

[46] Recall that the voting stage takes place at the beginning of the period. Subjects know which rule-sets have been in place in the two groups before the current voting stage, as well as the average contributions made by both groups in the previous period (see Section 3).

assigned. The average propensity to support sanctions also rises *(i)* for subjects located in groups that already implement sanctions before t, *(ii)* when the other group is also implementing sanctions, and *(iii)* when number of members in i's current group differs from 5. On the other hand, higher average contribution levels in the other group are associated with a decline in the probability that i votes for sanctions, but only when $Sanctions_{h \neq g,t-1} = 0$.

In order to gain further insights about the relationship between deviations from the optimal group size and the likelihood of voting for sanctions, column (2) splits the *abs(Group Size*$_{g,t}$ *- 5)* term into separate variables for deviations above and below 5. We find that $\mathbb{1}$(Group Size<5) (5 - Group Size$_{g,t}$) has a positive and significant marginal effect on $Pr(Vote_{i,t} = 1)$. By contrast, $\mathbb{1}$(Group Size>5) (Group Size$_{g,t}$ - 5) has no systematic influence on the probability of supporting sanctioning institutions. A possible interpretation for this result is that individuals located in groups facing the risk of depopulation vote for sanctions in order to retain or attract members (back).

Column (3), which adds a linear time trend to the column (2) specification, shows the probability of voting in favor of sanctions grows by more than 4.5 percentage points on average between (consecutive) voting stages, after controlling for other observed and unobserved factors. This finding reflects the patterns observed in the data and provides further support for Hypothesis 2.2, with the proportion of subjects voting for sanctions going from 41% in the first period in which a vote takes place to 62% in the last voting stage (see also Figure A.3 in the Appendix). Still, this leaves 38% of subjects voting against sanctioning institutions by the end of the experiment. There are at least three reasons that, either combined or in isolation, could help explain why such a large proportion of subjects keeps voting against sanctions. First, absent sanctions, contributions to the public good are still positive (see the left panel of Figure 3 and footnote 27). Second, although formal sanctions increase contributions, subjects do not invest the entirety of their endowment in the public good: as we showed in Section 5.1, the average contribution for groups with sanctions in VT is 46.61 tokens. Additionally, as suggested by Gürerk et al. (2006) and Sutter et al. (2010), individuals tend to dislike sanctioning institutions. Overall, the percentage of subjects voting for formal sanctions in the last voting stage is broadly aligned with the findings in Markussen et al. (2014).

The steady increase in the probability of voting for institutions over the course of the experiment is indicative of strong persistence in individual voting decisions. This result is confirmed in column (4), which adds the lagged dependent variable to the covariates of the column (3) specification. The probability of voting for sanctions rises by more than 45% for a subject who already voted in favor of sanctions in the previous voting stage. The marginal effect of $\mathbb{1}$(Group Size<5) (5 - Group Size$_{g,t}$) on $Pr(Vote_{i,t} = 1)$ ceases to be significant once we account for state dependence and for the growing propensity to support sanctions over time. The estimates for the other predictors, however, are robust to the inclusion of these controls.[47]

6 Conclusions

This paper analyzes the effect of the endogenous choice of sanctioning institutions on individuals' levels of cooperation and migration decisions. In our experiment, subjects are assigned to one of two groups and face a repeated social dilemma. Individuals can move between groups and, in the main treatment, they vote at fixed intervals on whether to implement a formal (centralized) sanctioning institution in the group in which they are located.

To study the effect of location choices on both individuals and societies, we introduce two features to our setting. First, subjects have group-specific preferences in that, all else equal, they enjoy higher payoffs if located in the group to which they are initially assigned. This feature accounts for the individual cost of migrating, which is generated by family and

[47] Table A.10 in the Appendix replicates the analyses in Table 5 accounting for covariate endogeneity through a control-function approach for random effects panel probit models (Giles and Murtazashvili, 2013). Unlike Fernández-Val and Vella (2011)'s method, this approach does not rely on large-T asymptotics, and is thus better suited for inference with only five periods. The estimates are in line with those in Table 5.

social ties. Second, to capture the cost imposed on societies by sizable migration outflows, we make the return from the public good (weakly) increasing in the group size. These two features imply, all else equal, an efficient partition of individuals across groups: at optimum, each individual contributes to the public good in the group to which she is initially assigned.

We compare the main treatment with one in which only one group is exogenously endowed with sanctioning institutions. When subjects cannot vote on the sanctioning regime, because sanctions positively affect contributions to the public good, the group endowed with sanctioning institutions (resp., the sanction-free group) tends to become overpopulated (resp., depopulated). Even though sanctions ensure high overall contributions to the public good, this migration flow causes payoff inequality both between groups and within the overpopulated group. By contrast, when subjects can vote on the sanctioning regime, groups increasingly implement sanctions over time, provide (more) efficient levels of the public good, and coordinate on the efficient partition of individuals across groups. Subjects' efficient location choices significantly increase overall payoffs and reduce payoff inequality between and within groups. At the individual level, subjects migrate less when voting over institutions is allowed; when migrating, they tend to move back to their initial group in order to cast their ballot and influence their initial group's institutional choice.

We conclude with three remarks. First, in our laboratory experiment, subjects manage to self-organize remarkably well in terms of group composition. Even though complex individual decisions – such as where to locate – can only be imperfectly approximated in a laboratory setting, this controlled environment allows us to take a close look at the multifaceted relationship between institutions, cooperation, and migration. This analysis, in turn, improves our understanding of groups' ability to select efficient outcomes. Second, our results seem to suggest formal institutions are sufficiently strong to stabilize migration patterns. In reality, informal sanctioning institutions – and their interaction with formal ones – also play a fundamental role in societies' ability to solve social dilemmas. Whether peer-to-peer sanctioning institutions also adapt to and stabilize migration patterns to achieve efficient outcomes remains an open question, which we leave for future research. Finally, this paper analyzes a framework in which the initial partition of subjects between groups coincides with the efficient one. Further research is required to study whether our main results hold in a different framework in which subjects' initial and efficient partitions do not coincide.

References

Acemoglu, D., Johnson, S., Robinson, J., 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. American Economic Review 91 (5), 1369–1401.

Acemoglu, D., Johnson, S., Robinson, J., 2005. The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth. American Economic Review 95 (3), 546–579.

Akashi, K., Kunitomo, K., 2012. Some properties of the LIML estimator in a dynamic panel structural equation. Journal of Econometrics 166 (2), 167–183.

Andreoni, J., Gee, L.K., 2012. Gun for hire: Delegated enforcement and peer punishment in public goods provision. Journal of Public Economics 96 (11), 1036–1046.

Angrist, J., 2001. Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice. Journal of Business & Economic Statistics 19 (1), 2–28.

Baldassarri, D., Grossman, G., 2011. Centralized sanctioning and legitimate authority promote cooperation in humans. Proceedings of the National Academy of Sciences 108 (27), 11023–11027.

Beine, M., Docquier, F., Rapoport, H., 2007. Measuring International Skilled Migration: A New Database Controlling for Age of Entry. The World Bank Economic Review 21 (2), 249–254.

Charness, G., Cobo-Reyes, R., Jimenez, N., 2014. Identities, selection, and contributions in a public-goods game. Games and Economic Behavior 87, 322–338.

Charness, G., Rabin, M., 2002. Understanding Social Preferences with Sample Tests. Quarterly Journal of Economics 117 (3), 817–869.

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

Conley, J., Konishi, H., 2002. Migration-proof Tiebout equilibrium: existence and asymptotic efficiency. Journal of Public Economics 86 (2), 243-262.

De Haas, H., 2010. Migration and development: A theoretical perspective. International Migration Review 44 (1), 227–264.

Denant-Boemont, L., Masclet, D., Noussair, C., 2007. Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. Economic Theory 33 (1), 145–167.

Dhaene, G., Jochmans, K., 2015. Split-panel Jackknife Estimation of Fixed-effect Models. Review of Economic Studies 83 (3), 991–1030.

Dinkelman, T., Schulhofer-Wohl, S., 2010. Migration, congestion externalities, and the evaluation of spatial investments. Journal of Development Economics 114, 189–202.

Dreber, A., Rand, D.G., Fudenberg, D., Nowak, M.A., 2008. Winners don't punish. Nature 452 (7185), 348–351.

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

Fehr, E., Gächter, S., 2000. Cooperation and Punishment in Public Goods Experiments. American Economic Review 90 (4), 980–994.

Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. Nature 415 (6868), 137–140.

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

Fehr, E., Williams, T., 2018. Social Norms, Endogenous Sorting and the Culture of Cooperation. Working Paper No. 267, Department of Economics, University of Zurich.

Fernández-Val, I., Vella, F., 2011. Bias corrections for two-step fixed effects panel data estimators. Journal of Econometrics 163 (2), 114–162.

Fernández-Val, I., Weidner, M., 2016. Individual and time effects in nonlinear panel models with large N, T. Journal of Econometrics 192 (1), 291–312.

Fréchette, G. (2012). Session-effects in the laboratory. Experimental Economics 15 (3), 485–498.

Gächter, S., Herrmann, B., 2009. Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. Philosophical Transactions of the Royal Society B: Biological Sciences 364 (1518), 791–806.

Giles, J., Murtazashvili, I., 2013. A Control Function Approach to Estimating Dynamic Probit Model with Endogenous Regressors. Journal of Econometric Methods 2 (1), 69–87.

Greenberg, J., Weber, S., 1986. Strong tiebout equilibrium under restricted preferences domain. Journal of Economic Theory 38 (1), 101–117.

Greene, W., 2012. Econometric Analysis. Essex: Pearson.

Gross, J., Méder, Z.Z., Okamoto-Barth, S., Riedl, A., 2016. Building the Leviathan – Voluntary centralisation of punishment power sustains cooperation in humans. Scientific Reports 6, 791–806.

Guala, F., 2012. Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. Behavioral and Brain Sciences 35 (1), 1–15.

Gürerk, O., Irlenbusch, B., Rockenbach, B., 2006. The Competitive Advantage of Sanctioning Institutions. Science 312 (5770), 108–111.

Gürerk, O., Irlenbusch, B., Rockenbach, B., 2009. Motivating teammates: The leader's choice between positive and negative incentives. Journal of Economic Psychology 4 (30), 591–607.

Gürerk, O., Irlenbusch, B., Rockenbach, B., 2014. On cooperation in open communities. Journal of Public Economics 120, 220–230.

Herrmann, B., Thöni, C., Gächter, S., 2008. Antisocial Punishment Across Societies. Science 319 (5868), 1362–1367.

Hill, J.K., 1987. Immigrant decisions concerning duration of stay and migratory frequency. Journal of Development Economics 25 (1), 221–234.

Honoré, B., 1992. Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects. Econometrica 60 (3), 221–234.

Hsiao, C., Zhou, Q., 2015. Statistical inference for panel dynamic simultaneous equations models. Journal of Econometrics 189 (2), 383–396.

Jones, E., 2003. The European Miracle: Environments, Economies and Geopolitics in the History of Europe and Asia. Cambridge: Cambridge University Press.

Katseli, L.T., Lucas, R.E., Xenogiani, T., 2006. Effects of Migration on Sending Countries: What do We Know? OECD Development Centre Working Paper No. 250.

Kosfeld, M., Okada, A., Riedl, A., 2009. Institution Formation in Public Goods Games. American Economic Review 99 (4), 1335–1355.

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

Marwell, G., Schmitt, D.R., 1972. Cooperation in a three-person Prisoner's Dilemma. Journal of Personality and Social Psychology 21 (3), 376–383.

Mayda, A.M., 2010. International migration: a panel data analysis of the determinants of bilateral flows. Journal of Population Economics 23 (4), 1249–1274.

Niedomysl, T., Amcoff, J., 2011. Why return migrants return: survey evidence on motives for internal return migration in Sweden. Population, Space and Place 17 (5), 656–673.

Nikiforakis, N., 2008. Punishment and counter-punishment in public good games: Can we really govern ourselves? Journal of Public Economics 92 (1), 91–112.

North, D.C., 1981. Structure and Change in Economic History. Cambridge: Cambridge University Press.

North, D.C., 1990. Institutions, Institutional Change and Economic Performance. Cambridge: Cambridge University Press.

North, D.C., Thomas, R.P., 1973. The Rise of the Western World: A New Economic History. Cambridge: Cambridge University Press.

Ostrom, E., Walker, J., Gardner, R., 1992. Covenants With and Without a Sword: Self-Governance Is Possible. American Political Science Review 86 (2), 404–417.

Pesarin, F., Salmaso, L., 2010. Permutation Tests for Complex Data: Theory, Applications and Software. Chichester: Wiley & Sons.

Poen, E., 2009. The Tobit model with feedback and random effects: A Monte-Carlo study. CeDEx Discussion Paper Series 2009-14, University of Nottingham.

Robbett, A., 2014. Local Institutions and the Dynamics of Community Sorting. American Economic Journal: Microeconomics 6 (3), 136–156.

Sahota, G.S., 1968. An Economic Analysis of Internal Migration in Brazil. Journal of Political Economy 76 (2), 218–245.

Sigmund, K., De Silva, H., Traulsen, A., Hauert, C., 2010. Social learning promotes institutions for governing the commons. Nature 466 (7308), 861–863.

Smith, A., 2013. Estimating the causal effect of beliefs on contributions in repeated public good games. Experimental Economics 16 (3), 414-425.

Stock, J., Yogo, M., 2005. Testing for Weak Instruments in Linear IV Regressions. In: Andrews, D., Stock, J. (Eds.). Identification and Inference for Econometric Models. New York: Cambridge University Press, 80–108.

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

Tiebout, C.M., 1956. A Pure Theory of Local Expenditures. Journal of Political Economy 64 (5), 416–424.

Traulsen, A., Röhl, T., Milinski, M., 2012. An economic experiment reveals that humans prefer pool punishment to maintain the commons. Proceedings of the Royal Society of London B: Biological Sciences 279 (1743), 3716–3721.

van der Heijden, E., Potters, J., Sefton, M., 2009. Hierarchy and opportunism in teams. Journal of Economic Behavior & Organization 69 (1), 39–50.

Wooldridge, J., 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of Applied Econometrics 20 (1), 39–54.

Wooldridge, J., 2010. Econometric Analysis of Cross Section and Panel Data. Cambridge: MIT Press.

Yamagishi, T., 1986. The provision of a sanctioning system as a public good. Journal of Personality and Social Psychology 51 (1), 110–116.

Zhang, B., Li, C., De Silva, H., Bednarik, P., Sigmund, K., 2014. The evolution of sanctioning institutions: an experimental approach to the social contract. Experimental Economics 17 (2), 285–303.

Appendix

A. Experimental Instructions

A.1. First Activity

You will be in a group of 5 people. You will receive 50 tokens (converted to cash at the end of the experiment) and will choose to allocate some portion of these to your own individual account and some portion to the group account. Your final payoffs will be as follows:

$$\boxed{\begin{array}{l} 50 - (\text{tokens you allocate to group account}) + (1/5) \times 2.250\times \\ (\text{sum of tokens allocated by all in group to group account}) \end{array}}$$

At the end of the experiment tokens will be converted into GBP at the rate of 30 tokens = 1 GBP.

A.2. Second Activity: *Voting* Treatment

There will be a total of 30 periods in the second activity. We will carefully explain what you have to do in this second part.

Allocations: In each period, you will be in a group of some size (you could be also by yourself). In each period, you will receive 50 tokens (converted to cash at the end of the experiment) and will choose to allocate some portion of these to your own individual account and some portion to the group account. Your final payoffs will be as follows:

$$\boxed{\begin{array}{l} 50 - (\text{tokens you allocate to group account}) + (1/n)\times \text{ Factor } \times \\ (\text{sum of tokens allocated by all in group to group account}) \end{array}}$$

where n is the number of members of the group. The Factor that multiplies the sum of the tokens is explained below.

Groups: In the first period you will be assigned to a five people group (including yourself). You can be assigned to either GROUP A or GROUP B. At the end of each period you will have the chance of deciding whether you want to leave your group. We will explain in more detail how you do this.

Group size and multiplying Factor: The total number of tokens that end up in the group account is multiplied by a Factor. The table below shows the values of the Factor depending on: *(i)* the group you are in and *(ii)* the group size.

Group Size (n)	**Factor if in the Initial Group**	**Factor if <u>not</u> in the Initial Group**
1	1.00	1.00
2	1.50	1.15
3	1.75	1.40
4	2.00	1.65
5	2.25	1.90
6	2.25	1.90
7	2.25	1.90
8	2.25	1.90
9	2.25	1.90
10	2.25	1.90

Each token allocated to the group account yields a return higher than one unit and all members of the group share this return equally.

Note that the multiplying factor increases as the group size increases. The increment of the factor with the group size will be capped at 5 members in the group. If the size of the group goes above that, the factor will remain the same as if the size of the group was 5.

There will be two different factors depending on whether you are in your **initial** group or you decided to switch groups. Given the size of the group, the factor will be always larger if you are in your **initial** group than if you are in the other group. Thus, for the same amount of tokens allocated to the group account, and for the same size of the group, tokens allocated to the group account will generate more money if you are in your **initial** group.

Each token allocated to one's private account pays one unit to that individual, independently of the group size.

EXAMPLE: Consider the case in which one participant belongs to **Group A** and her initial group was **Group A**. If in a five-person group the total number of tokens in the group account is 100, that participant receives a payoff of $(100 \times 2.250)/5=45$ tokens from the group account (plus the number of tokens allocated to her private account). By comparison, suppose the participant is in **Group B** and her initial group was **Group A**. If in a five-person group the total number of tokens in the group account is 100, then, that participant receives a payoff of $(100 \times 1.900)/5=38$ tokens from the group account (plus the number of tokens allocated to her private account).

Payment rules: there will be **two** different **Rule Sets**, which affect your earnings in different ways:

RULE SET 1 (no points reduction): In Rule Set 1, earnings are determined in exactly the same way as explained in the **Allocations** section. So, payoffs will be computed as follows

50 – (points you allocate to group account) + $(1/n)\times$ Factor $\times$ (sum of points allocated by all in group to group account)

RULE SET 2 (automatic points reduction): In Rule Set 2, each individual pays a fixed fee of 15 tokens in each period. The fee is deducted from your earnings at the end of the period. In addition, each individual pays a fine equal to 80 percent of the amount of tokens allocated to the private account in that period. Payoffs in each period are calculated as follows:

50 – (points you allocate to group account) + $(1/n)\times$ Factor $\times$ (sum of points allocated by all in group to group account) – 0.8 $\times$(points you allocate to private account) – 15

EXAMPLE: Consider the case in which one participant belongs to a group of 5 people and RULE SET 2 (automatic points reduction) is in place. If the participant contributes 20 tokens to the group account and keeps 30 tokens for the private account, given RULE SET 2, that participant would have to pay a fine of $0.8 \times 30{=}24$ tokens and then, she would keep 6 tokens out of 30.

Payment rules in first 5 periods: If you were assigned to Group A, Rule Set 1 will apply for the first 5 periods. If you were assigned to group B, Rule set 2 will apply for the first 5 periods.

Group determination: At the end of each period, you will observe: *(i)* the average contribution of your current group, *(ii)* the average contribution of the other group, *(iii)* the rule set of your current group, and *(iv)* the rule set of the other group. Then, you will decide whether you wish to leave the group and move to the other one.

Voting for rules: Initial rules for Groups A and B will apply only for the first 5 periods. Every five periods you will have the chance to vote on whether you want Rule Set 1 or Rule Set 2 to be applied to the group you currently belong to. Note, then, that when you vote, you will vote to establish a rule in the group that you belong to in the voting period.

Payment: At the end of the experiment tokens will be converted into GBP at the rate of 30 tokens = 1GBP. You will be paid only for 3 randomly chosen periods.

SUMMARY

1. There will be a total of 30 periods.
2. You will begin by being in a five-people group.
3. In each period, you will allocate 50 tokens between a private and a group account.
4. The returns from the group account depend on the size of the group and on whether you are participating in your initial group or not.
5. You can decide at the end of the period whether you want to switch groups or not.
6. There are two payment rules. Every five period you will vote which of the two rules you prefer.
7. You will be paid only for 3 rounds (randomly chosen).

A.3. Second Activity: *No-Voting* Treatment

There will be a total of 30 periods in the second activity. We will carefully explain what you have to do in this second part.

Allocations: In each period, you will be in a group of some size (you could be also by yourself). In each period, you will receive 50 tokens (converted to cash at the end of the experiment) and will choose to allocate some portion of these to your own individual account and some portion to the group account. Your final payoffs will be as follows:

$$\boxed{\begin{array}{l} 50 - (\text{tokens you allocate to group account}) + (1/n)\times \text{Factor} \times \\ (\text{sum of tokens allocated by all in group to group account}) \end{array}}$$

where n is the number of members of the group. The Factor that multiplies the sum of the tokens is explained below.

Groups: In the first period you will be assigned to a five people group (including yourself). You can be assigned to either GROUP A or GROUP B. At the end of each period you will have the chance of deciding whether you want to leave your group. We will explain in more detail how you do this.

Group size and multiplying Factor: The total number of tokens that end up in the group account is multiplied by a Factor. The table below shows the values of the Factor depending on: *(i)* the group you are in and *(ii)* the group size.

Group Size (n)	**Factor if in the Initial Group**	**Factor if not in the Initial Group**
1	1.00	1.00
2	1.50	1.15
3	1.75	1.40
4	2.00	1.65
5	2.25	1.90
6	2.25	1.90
7	2.25	1.90
8	2.25	1.90
9	2.25	1.90
10	2.25	1.90

Each token allocated to the group account yields a return higher than one unit and all members of the group share this return equally.

Note that the multiplying factor increases as the group size increases. The increment of the factor with the group size will be capped at 5 members in the group. If the size of the group goes above that, the factor will remain the same as if the size of the group was 5.

There will be two different factors depending on whether you are in your **initial** group or you decided to switch groups. Given the size of the group, the factor will be always larger if you are in your **initial** group than if you are in the other group. Thus, for the same amount of tokens allocated to the group account, and for the same size of the group, tokens allocated to the group account will generate more money if you are in your **initial** group.

Each token allocated to one's private account pays one unit to that individual, independently of the group size.

EXAMPLE: Consider the case in which one participant belongs to **Group A** and her initial group was **Group A**. If in a five-person group the total number of tokens in the group account is 100, that participant receives a payoff of $(100 \times 2.250)/5{=}45$ tokens from the group account (plus the number of tokens allocated to her private account). By comparison, suppose the participant is in **Group B** and her initial group was **Group A**. If in a five-person group the total number of tokens in the group account is 100, then, that participant receives a payoff of $(100 \times 1.900)/5{=}38$ tokens from the group account (plus the number of tokens allocated to her private account).

Payment rules: there will be **two** different **Rule Sets**, which affect your earnings in different ways:

RULE SET 1 (no points reduction): In Rule Set 1, earnings are determined in exactly the same way as explained in the **Allocations** section. So, payoffs will be computed as follows

50 – (points you allocate to group account) $+ (1/n)\times$ Factor $\times$ (sum of points allocated by all in group to group account)

RULE SET 2 (automatic points reduction): In Rule Set 2, each individual pays a fixed fee of 15 tokens in each period. The fee is deducted from your earnings at the end of the period. In addition, each individual pays a fine equal to 80 percent of the amount of tokens allocated to the private account in that period. Payoffs in each period are calculated as follows:

50 – (points you allocate to group account) + $(1/n)\times$ Factor $\times$ (sum of points allocated by all in group to group account) – 0.8 $\times$(points you allocate to private account) – 15

EXAMPLE: Consider the case in which one participant belongs to a group of 5 people and RULE SET 2 (automatic points reduction) is in place. If the participant contributes 20 tokens to the group account and keeps 30 tokens for the private account, given RULE SET 2, that participant would have to pay a fine of $0.8 \times 30=24$ tokens and then, she would keep 6 tokens out of 30.

Payment rules: If you were assigned to Group A, Rule Set 1 will apply. If you were assigned to group B, Rule set 2 will apply.

Group determination: At the end of each period, you will observe: *(i)* the average contribution of your current group, *(ii)* the average contribution of the other group, *(iii)* the rule set of your current group, and *(iv)* the rule set of the other group. Then, you will decide whether you wish to leave the group and move to the other one.

Payment: At the end of the experiment tokens will be converted into GBP at the rate of 30 tokens = 1GBP. You will be paid only for 3 randomly chosen periods.

SUMMARY

1. There will be a total of 30 periods.
2. You will begin by being in a five-people group.
3. In each period, you will allocate 50 tokens between a private and a group account.
4. The returns from the group account depend on the size of the group and on whether you are participating in your initial group or not.

5. You can decide at the end of the period whether you want to switch groups or not.

6. There are two payment rules.

7. You will be paid only for 3 rounds (randomly chosen).

B. Additional Results

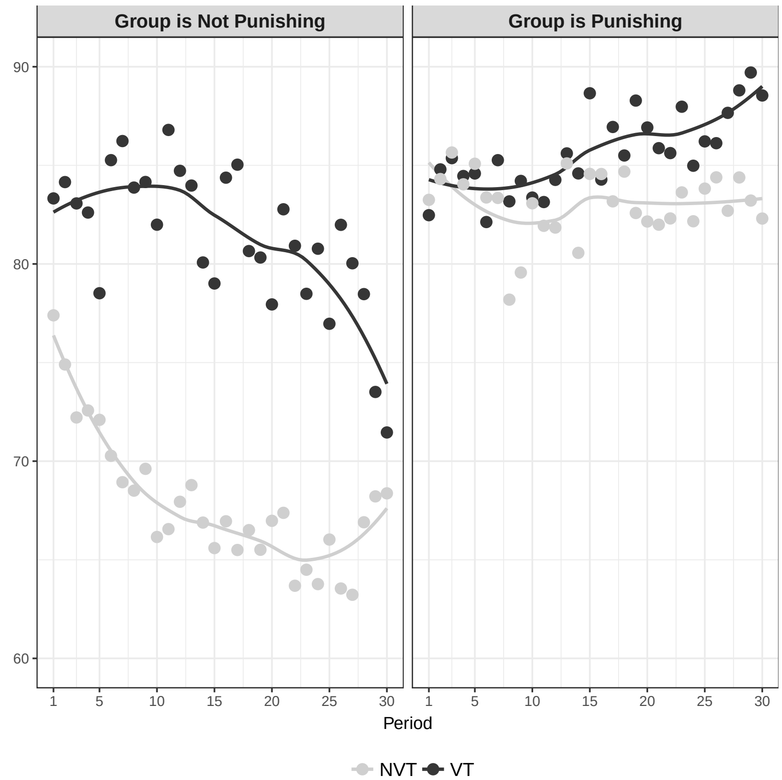

Figure A.1: Average participant payoff per period, by institutional setting and treatment. The left (right) panel plots the average participant payoff per period in groups without (with) sanctioning institutions. In both cases, black (gray) dots represent the mean payoff per period - averaged across sessions - under VT (NVT); solid lines represent locally weighted scatterplot smoothing curves fitted to the data. As in the analysis of average contributions (Figure 3), we find that, in both NVT and VT, the average per-period payoffs are significantly higher for sanctioning than for sanction-free groups (Wilcoxon signed-rank tests: z=4.34, p=0.00 for NVT, and z=3.51, p=0.00 for VT, two-tailed). In groups implementing sanctions, the average payoffs per period are not significantly different across treatments (Mann-Whitney test: z=1.54, p=0.13, two-tailed). By contrast, in sanction-free groups, the average payoff per period is larger in VT than in NVT (Mann-Whitney test: z=3.34, p=0.00, two-tailed).

Table A.1
Parameter Estimates - Censored Regression Models for Contributions

	(1)	(2)	(3)	(4)	(5)
Intercept	18.84***	18.93***	15.15***	18.94***	18.82***
	(1.10)	(1.13)	(1.10)	(1.82)	(1.72)
$Sanctions_{g,t}$	33.19***	33.10***	31.13***	26.24***	25.82***
	(0.77)	(0.76)	(0.74)	(3.12)	(2.87)
$Initial\ Group_{i,t}$	4.14***	4.25***	3.28***	1.23*	0.89
	(0.75)	(0.78)	(0.74)	(0.74)	(0.75)
$Group\ Contribution_{g[-i],t-1}$	0.30***	0.30***	0.08***	-0.03	-0.02
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
$abs(Group\ Size_{g,t}$ - 5)	-0.86***		-0.70***	0.95***	
	(0.24)		(0.22)	(0.25)	
$\mathbb{1}(Group\ Size_{g,t}<5)(5 - Group\ Size_{g,t})$		-1.13***			2.45***
		(0.39)			(0.39)
$\mathbb{1}(Group\ Size_{g,t}>5)(Group\ Size_{g,t} - 5)$		-0.78***			0.27
		(0.25)			(0.28)
$Contribution_{i,t-1}$			0.35***	0.28***	0.28***
			(0.02)	(0.02)	(0.02)
Observations	7,250	7,250	7,250	7,250	7,250
Log likelihood	-24,993.46	-24,987.21	-23,777.47	-24,246.43	-24,215.80

Notes. The table reports "raw" parameter estimates for the panel data models used to compute the marginal effects reported in Table 3. Columns (1)-(3) are fitted to data from the *Voting* treatment, whereas columns (4) and (5) use data from the *No-Voting* condition. All the specifications include subject, period, group, and session random effects. Standard errors are reported in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%.

Table A.2
Marginal Effects of Covariates on Individual Contributions
Estimated from Fixed-Effects Censored Regression Models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Sanctions_{g,t}$	14.77***	14.74***	13.42***	14.23***	14.90***	19.46***	19.29***
	(0.30)	(0.30)	(0.29)	(0.33)	(0.41)	(1.10)	(1.09)
$Initial\ Group_{i,t}$	1.41**	1.53**	1.24***	0.19	0.06	1.77**	2.13**
	(0.33)	(0.35)	(0.32)	(0.33)	(0.34)	(0.81)	(0.82)
$Group\ Contribution_{g[-i],t-1}$	0.13***	0.13***	0.03**	-0.02*	-0.02*	0.15***	0.14***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
$abs(Group\ Size_{g,t}$ - 5)	-0.46***		-0.42***	0.37***		-0.42***	
	(0.10)		(0.10)	(0.10)		(0.14)	
$\mathbb{1}(Group\ Size_{g,t} < 5)(5 - Group\ Size_{g,t})$		-0.58***			0.79***		-1.07***
		(0.18)			(0.17)		(0.34)
$\mathbb{1}(Group\ Size_{g,t} > 5)(Group\ Size_{g,t} - 5)$		-0.44**			0.21		-0.31**
		(0.10)			(0.13)		(0.15)
$Contribution_{i,t-1}$			0.16***	0.14***	0.14***		
			(0.01)	(0.01)	(0.01)		
Observations	7,250	7,250	7,250	7,250	7,250	7,250	7,250

Notes. The table reports marginal effects of the covariates on subjects' contributions, obtained from fixed-effects panel data censored regression models. Columns (1)-(5) replicate the analyses in columns (1)-(5) of Table 3, using the split-panel jackknife estimation approach proposed by Dhaene and Jochmans (2015) to correct for incidental parameter bias. For robustness, columns (6) and (7) replicate the analyses in columns (1) and (2), respectively, using Honoré (1992)'s semiparametric estimator (which cannot accommodate lagged dependent variables). Columns (1)-(3) and (6)-(7) are fitted to data from the *Voting* treatment, while columns (4) and (5) use data from the *No-Voting* condition. All specifications include subject fixed-effects. Standard errors are reported in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%.

Table A.3
Determinants of Individual Contributions
Model Fitted to Pooled Data from the *Voting* and *No-Voting* treatments

	(1)	(2)
Intercept	16.89***	
	(0.90)	
$Sanctions_{g,t}$	30.90***	12.74***
	(1.40)	(0.58)
$Sanctions_{g,t} \times VT$	2.25	0.93
	(1.54)	(0.63)
$Initial\ Group_{i,t}$	1.07	0.44
	(0.67)	(0.28)
$Initial\ Group_{i,t} \times VT$	1.97**	0.81**
	(0.96)	(0.39)
$Group\ Contribution_{g[-i],t-1}$	-0.03	-0.01
	(0.02)	(0.01)
$Group\ Contribution_{g[-i],t-1} \times VT$	0.07**	0.02**
	(0.03)	(0.01)
$abs(Group\ Size_{g,t}$ - 5)	0.92***	0.38***
	(0.23)	(0.10)
$abs(Group\ Size_{g,t}$ - 5) $\times VT$	-1.90***	-0.78***
	(0.32)	(0.13)
$Contribution_{i,t-1}$	0.27***	0.11***
	(0.02)	(0.01)
$Contribution_{i,t-1} \times VT$	0.07**	0.03***
	(0.03)	(0.01)
Observations	14,500	
Log likelihood	-48,150.42	

Notes. The table reports parameter estimates (column 1) and marginal effects (column 2) for a model of individual contributions fitted to pooled data from the *Voting* and *No-Voting* treatments. The specification includes the same explanatory variables as those in columns (3) and (4) of Table 3, adding interactions between the predictors and an indicator for the *Voting* treatment. Subject, period, group, and session random effects are incorporated in the model specification as well. Standard errors are presented in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%.

Table A.4
Determinants of Individual Contributions - Accounting for Covariate Endogeneity Using Fernández-Val and Vella (2011)'s Two-Step Estimator

	(1)	(2)	(3)	(4)	(5)
$Sanctions_{g,t}$	12.91***	12.89***	10.92***	14.43***	14.57***
	(1.41)	(1.67)	(2.47)	(0.57)	(0.66)
$Initial\ Group_{i,t}$	2.72**	2.76**	2.33*	4.57***	1.34
	(1.25)	(1.30)	(1.35)	(0.38)	(0.87)
$Group\ Contribution_{g[-i],t-1}$	0.16***	0.16***	0.06***	-0.01	-0.01
	(0.02)	(0.03)	(0.02)	(0.01)	(0.01)
$abs(Group\ Size_{g,t}$ - 5)	-0.46		-0.48	0.31***	
	(0.39)		(0.39)	(0.12)	
$\mathbb{1}(Group\ Size_{g,t}<5)(5 - Group\ Size_{g,t})$		-0.45			0.76***
		(0.40)			(0.14)
$\mathbb{1}(Group\ Size_{g,t}>5)(Group\ Size_{g,t} - 5)$		-0.44			0.20
		(0.46)			(0.13)
$Contribution_{i,t-1}$			0.17***	0.16***	0.15***
			(0.03)	(0.01)	(0.01)
Observations	7,250	7,250	7,250	7,250	7,250

Notes. The table replicates the analyses in Table 3, reporting marginal effects of the covariates on subjects' contributions estimated using Fernández-Val and Vella (2011)'s two-step (control function) approach for non-linear panel data models. Columns (1)-(4) are fitted to data from the *Voting* treatment, while columns (4) and (5) use data from the *No-Voting* condition. These specifications treat $Sanctions_{g,t}$ (in VT), $Initial\ Group_{i,t}$ and the measures of deviations from the optimal group size as endogenous, and allow for dynamic feedback from the dependent variable to the predictors. The control variables are the (generalized) residuals from the reduced-form models for the endogenous variables; the first-step estimates are available from the authors upon request. All specifications in the table include subject fixed-effects; although time dummies (or time trends) are not covered by Fernández-Val and Vella (2011)'s regularity conditions, adding them does not alter the results. Standard errors are reported in parentheses; we combined split-panel jackknife estimation and (panel) bootstrapping to correct for incidental parameter bias and adjust for generated regressors (Dhaene and Jochmans, 2015). Significance levels: *** at 1%, ** at 5%, * at 10%.

Table A.5

Parameter Estimates - Probit Panel Models for Migration

	(1)	(2)	(3)	(4)	(5)
Intercept	-1.27***	-1.31***	-1.36***	-1.39***	-0.26
	(0.24)	(0.24)	(0.25)	(0.25)	(0.19)
$Sanctions_{g,t}$	-0.52***	-0.52***	-0.58***	-0.55***	-1.00***
	(0.06)	(0.06)	(0.07)	(0.07)	(0.11)
$Initial\ Group_{i,t}$	-1.12***	-1.12***	-1.10***	-1.08***	-0.82***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
$Contribution_{i,t}$ - $Group\ Contribution_{g[-i],t}$	0.21***	0.21***	0.21***	0.21***	0.10***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
$abs(Group\ Size_{g,t}$ - 5)	0.14***	0.14***	0.15***	0.15***	0.06***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$Sanctions_{h \neq g,t}$	0.25***	0.25***	0.25***	0.25***	
	(0.07)	(0.07)	(0.07)	(0.07)	
$Group\ Contribution_{h \neq g,t}$	0.01	0.01	0.01	0.02	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
$Vote\ Different\ from\ the\ Group_{i,t-1}$		0.36***	0.20	0.20	
		(0.10)	(0.14)	(0.14)	
$Vote\ Different\ from\ the\ Group_{i,t-1} \times Sanctions_{g,t}$			0.30	0.32	
			(0.19)	(0.19)	
$Period\ Before\ a\ Voting\ Stage_t$				0.24**	
				(0.11)	
$Period\ Before\ a\ Voting\ Stage_t \times Initial\ Group_{i,t}$				-0.31**	
				(0.14)	
Observations	7,250	7,250	7,250	7,250	7,250
Log likelihood	-2,079.81	-2,070.96	-2,069.19	-2,066.51	-2,568.14

Notes. The table reports "raw" parameter estimates for the panel probit models used to compute the marginal effects reported in Table 4. Columns (1)-(4) are fitted to data from the *Voting* treatment. Column (5) uses data from the *No-Voting* condition; in this specification, $Sanctions_{h \neq g,t}$ is dropped from the predictors due to its perfect collinearity with $Sanctions_{g,t}$. All the models include subject, period, group, and session random effects. Standard errors are reported in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%.

Table A.6

Parameter Estimates - Fixed-Effects Probit Panel Models for Migration

	(1)	(2)	(3)	(4)	(5)
$Sanctions_{g,t}$	-0.55***	-0.55***	-0.57***	-0.58***	-0.74***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)
$Initial\ Group_{i,t}$	-0.97***	-0.97***	-0.96***	-0.92***	-0.61***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
$Contribution_{i,t}$ - $Group\ Contribution_{g[-i],t}$	0.20***	0.20***	0.20***	0.20***	0.09***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
$abs(Group\ Size_{g,t}$ - $5)$	0.16***	0.16***	0.16***	0.16***	0.06***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
$Sanctions_{h \neq g,t}$	0.28***	0.28***	0.27***	0.27***	
	(0.07)	(0.07)	(0.07)	(0.07)	
$Group\ Contribution_{h \neq g,t}$	0.06**	0.06**	0.06**	0.06**	0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
$Vote\ Different\ from\ the\ Group_{i,t-1}$		0.38***	0.19	0.18	
		(0.12)	(0.16)	(0.16)	
$Vote\ Different\ from\ the\ Group_{i,t-1} \times Sanctions_{g,t}$			0.35*	0.37*	
			(0.20)	(0.20)	
$Period\ Before\ a\ Voting\ Stage_t$				0.20*	
				(0.11)	
$Period\ Before\ a\ Voting\ Stage_t \times Initial\ Group_{i,t}$				-0.30**	
				(0.14)	
Observations	5,336	5,336	5,336	5,336	5,626
Log likelihood	-1,841.11	-1,836.05	-1,834.20	-1,831.55	-2,306.97

Notes. The table reports parameter estimates from bias-corrected fixed-effects panel probit models for migration using Fernández-Val and Weidner (2016)'s method. The bias correction is obtained from jackknife estimates; using an analytical correction yields virtually identical results. Columns (1)-(4) are fitted to data from the *Voting* treatment. Column (5) uses data from the *No-Voting* condition; in this specification, $Sanctions_{h \neq g,t}$ is dropped from the predictors due to its perfect collinearity with $Sanctions_{g,t}$. Differences in the number of observations vis-à-vis Tables 4 and A.5 are due to the fact that Fernández-Val and Weidner (2016)'s method drops subjects for whom the dependent variable does not change over time. All the models include subject fixed-effects. Columns (1)-(3) and (5) also include period fixed-effects, which must be dropped in column (4) in order to recover the coefficient for $Period\ Before\ a\ Voting\ Stage_t$. Standard errors are reported in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%.

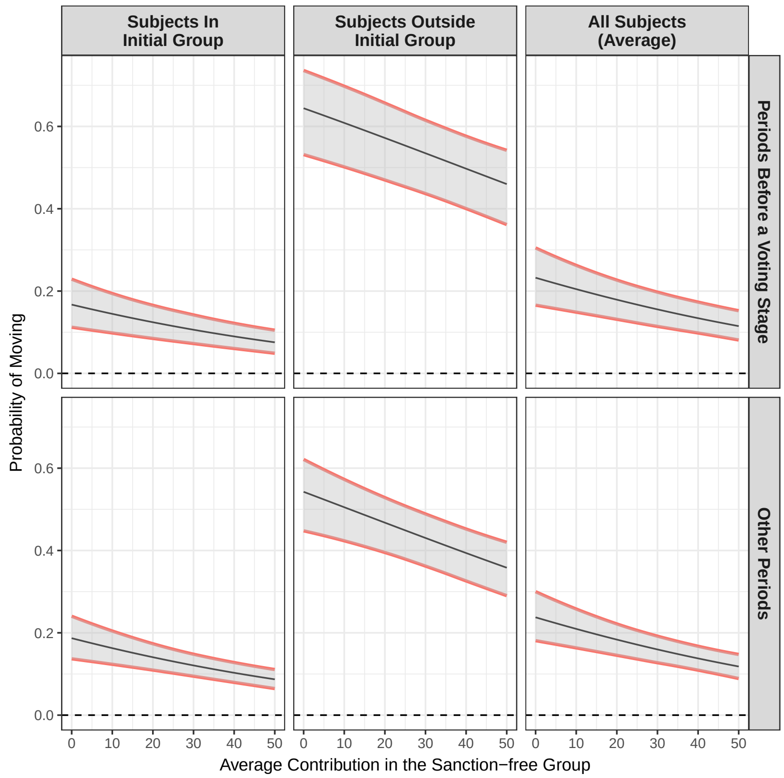

Figure A.2: Probability of moving from sanction-free to sanctioning groups, as a function of the average contribution levels in the group without sanctions. The left (middle) column presents the probability of moving from a group without sanctions to a group with sanctions when subjects are located in (outside) their initial group; the right column displays the average probability for all subjects. The upper panel reports $Pr(Migration_{i,t} = 1)$ right before a voting stage; the lower panel reports the probability of moving in periods other than those immediately preceding a voting stage. Solid lines represent point estimates obtained from a model similar to that in column (4) of Table 4, replacing $Contribution_{i,t}$ - $Group\ Contribution_{g[-i],t}$ with $Group\ Contribution_{g,t}$. Shaded areas give the 95% confidence intervals.

Table A.7
Determinants of Individual Migration Decisions - Accounting for Covariate Endogeneity using Fernández-Val and Vella (2011)'s Two-Step Estimator

	(1)	(2)	(3)	(4)	(5)
$Sanctions_{g,t}$	-10.04***	-9.98***	-9.90***	-9.80***	-17.29***
	(2.52)	(2.53)	(2.50)	(2.50)	(2.02)
$Initial\ Group_{i,t}$	-13.65***	-13.66***	-13.58***	-12.43***	-8.78*
	(3.86)	(3.85)	(3.87)	(3.87)	(4.51)
$Contribution_{i,t}$ - $Group\ Contribution_{g[-i],t}$	2.96***	2.97***	2.99***	2.95***	1.67***
	(0.65)	(0.63)	(0.63)	(0.64)	(0.64)
$abs(Group\ Size_{g,t}$ - 5)	1.37	1.41	1.40	1.44	-0.79
	(0.94)	(0.94)	(0.94)	(0.96)	(2.21)
$Sanctions_{h\neq g,t}$	2.88**	2.68**	2.67**	2.73**	
	(1.37)	(1.36)	(1.36)	(1.35)	
$Group\ Contribution_{h\neq g,t}$	0.09	0.12	0.11	0.13	0.04
	(0.08)	(0.08)	(0.09)	(0.09)	(0.10)
$Vote\ Different\ from\ the\ Group_{i,t-1}$		6.55**			
		(2.65)			
$Vote\ Different\ from\ the\ Group_{i,t-1}$ *under No Sanctions*			4.51	4.38	
			(4.37)	(4.36)	
$Voted\ Against\ the\ Group_{i,t-1}$ *under Sanctions*			7.61**	7.86**	
			(3.56)	(3.58)	
$Period\ Before\ a\ Voting\ Stage_t$ - $Initial\ Group_{i,t}$				-1.22	
				(1.67)	
$Period\ Before\ a\ Voting\ Stage_t$ - $Outside\ Initial\ Group_{i,t}$				5.63**	
				(2.76)	
Observations	7,250	7,250	7,250	7,250	7,250

Notes. The table replicates the analyses in Table 4, reporting marginal effects of the covariates on the probability of moving (in percentage points), using Fernández-Val and Vella (2011)'s two-step (control function) approach for non-linear panel data models. Columns (1)-(4) are fitted to data from the *Voting* treatment, while column (5) uses data from the *No-Voting* condition. These specifications treat $Sanctions_{g,t}$ (in *VT*), $Initial\ Group_{i,t}$ and $abs(Group\ Size_{g,t}$ - 5) as endogenous, and allow for dynamic feedback from the dependent variable to the predictors. The control variables are the (generalized) residuals from the reduced-form models for the endogenous variables; the first-step estimates are available from the authors upon request. All specifications include subject fixed-effects; although time dummies (or time trends) are not covered by Fernández-Val and Vella (2011)' regularity conditions, adding them does not alter the results. Standard errors are reported in parentheses; we combined split-panel jackknife estimation and (panel) bootstrapping to correct for incidental parameter bias and adjust for generated regressors (Dhaene and Jochmans, 2015). Significance levels: *** at 1%, ** at 5%, * at 10%.

Table A.8
Parameter Estimates - Probit Panel Models for Voting

	(1)	(2)	(3)	(4)
Intercept	-1.13***	-1.12***	-2.02***	-1.42***
	(0.29)	(0.29)	(0.38)	(0.25)
$Sanctions_{g,t-1}$	0.65***	0.63***	0.56***	0.50***
	(0.16)	(0.16)	(0.16)	(0.11)
$Initial\ Group_{i,t}$	0.51***	0.53***	0.48***	0.30**
	(0.16)	(0.17)	(0.16)	(0.13)
$Contribution_{i,t-1}$ - $Group\ Contribution_{g[-i],t-1}$	-0.08	-0.08	-0.08	0.01
	(0.06)	(0.06)	(0.06)	(0.05)
$abs(Group\ Size_{g,t}$ - $5)$	0.11*			
	(0.06)			
$Sanctions_{h \neq g,t-1}$	0.78**	0.77**	0.75**	0.69***
	(0.34)	(0.34)	(0.34)	(0.24)
$Group\ Contribution_{h \neq g,t-1}$	-0.27***	-0.26**	-0.23**	-0.25***
	(0.09)	(0.09)	(0.09)	(0.07)
$Sanctions_{h \neq g,t-1} \times Group\ Contribution_{h \neq g,t-1}$	-0.25	-0.26	-0.36	-0.35
	(0.38)	(0.39)	(0.37)	(0.29)
$\mathbb{1}(Group\ Size_{g,t}<5)(5 - Group\ Size_{g,t})$		0.13**	0.12**	0.04
		(0.06)	(0.06)	(0.05)
$\mathbb{1}(Group\ Size_{g,t}>5)(Group\ Size_{g,t} - 5)$		0.06	0.04	0.03
		(0.09)	(0.08)	(0.07)
Time Trend			0.04***	0.03***
			(0.01)	(0.01)
$Vote_{i,t-1}$				1.33***
				(0.10)
Observations	1,250	1,250	1,250	1,000
Log likelihood	-648.06	-647.78	-641.67	-488.31

Notes. The table reports "raw" parameter estimates for the panel probit models used to compute the marginal effects reported in Table 5. All the columns report estimates obtained under the *Voting* treatment (subjects cannot vote in *NVT*). All specifications include subject, period, group, and session random effects; columns (3)-(4) also include a linear time trend. Standard errors are reported in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%.

Table A.9
Parameter Estimates
Fixed-Effects Probit Panel Models for Voting

	(1)	(2)	(3)	(4)
$Sanctions_{g,t-1}$	0.35**	0.33**	0.05	-0.05
	(0.14)	(0.14)	(0.15)	(0.24)
$Initial\ Group_{i,t}$	0.37**	0.40**	0.39**	0.28
	(0.18)	(0.18)	(0.19)	(0.24)
$Contribution_{i,t-1}$ - $Group\ Contribution_{g[-i],t-1}$	-0.07	-0.07	-0.06	-0.10
	(0.06)	(0.07)	(0.07)	(0.09)
$abs(Group\ Size_{g,t}$ - $5)$	0.12**			
	(0.06)			
$Sanctions_{h \neq g,t-1}$	-0.05	-0.03	-0.22	0.47**
	(0.15)	(0.15)	(0.17)	(0.23)
$Group\ Contribution_{h \neq g,t-1}$	-0.07	-0.07	-0.15	-0.28**
	(0.10)	(0.10)	(0.11)	(0.14)
$Sanctions_{h \neq g,t-1} \times Group\ Contribution_{h \neq g,t-1}$	0.05	0.06	0.06	0.11
	(0.11)	(0.11)	(0.11)	(0.15)
$\mathbb{1}(Group\ Size_{g,t}<5)(5 - Group\ Size_{g,t})$		0.14**	0.06	0.01
		(0.07)	(0.07)	(0.09)
$\mathbb{1}(Group\ Size_{g,t}>5)(Group\ Size_{g,t} - 5)$		0.06	0.03	0.01
		(0.09)	(0.10)	(0.12)
Time Trend			0.05***	0.06***
			(0.01)	(0.01)
$Vote_{i,t-1}$				0.37**
				(0.18)
Observations	665	665	665	420
Log likelihood	-374.63	-374.04	-345.67	-222.67

Notes. The table reports parameter estimates from bias-corrected fixed-effects panel probit models for $Vote_{i,t}$ using Fernández-Val and Weidner (2016)'s method. The bias correction is obtained from jackknife estimates; using an analytical correction yields virtually identical results. Note that Fernández-Val and Weidner (2016)'s method relies on large-T (and large-N) asymptotics, and thus is not particularly well suited for the analysis of data from only five periods (those including a voting stage). Differences in the number of observations vis-à-vis Tables 5 and A.8 are due to the fact that Fernández-Val and Weidner (2016)'s method drops subjects for whom the dependent variable does not change over time. All the models include subject fixed-effects. Columns (1) and (2) also include period fixed-effects, which must be dropped in columns (3) and (4) due to the addition of a *Time Trend* to the model specification. All the columns report estimates obtained under the *Voting* treatment (subjects cannot vote in NVT). Standard errors are reported in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%.

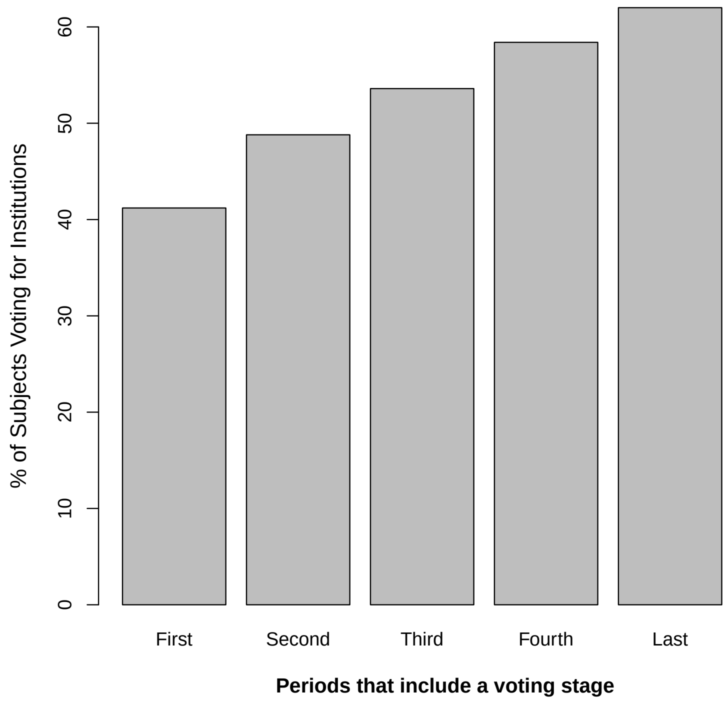

Figure A.3: Proportion of subjects voting for sanctions over time. Bars represent the proportion of subjects in VT voting for institutions in each of the periods in which a voting stage takes place, averaged across sessions.

Table A.10
Determinants of Individual Voting Decisions
Accounting for Covariate Endogeneity Using
Giles and Murtazashvili (2013)'s Control Function Approach

	(1)	(2)	(3)	(4)
$Sanctions_{g,t-1}$	11.35***	11.02***	10.61**	11.12**
	(4.12)	(4.18)	(4.20)	(4.97)
$Initial\ Group_{i,t}$	10.27***	10.57***	11.03***	8.09*
	(3.65)	(3.82)	(4.03)	(4.46)
$Contribution_{i,t-1}$ - $Group\ Contribution_{g[-i],t-1}$	-0.03	0.16	0.04	0.28
	(1.72)	(1.79)	(1.88)	(2.01)
$abs(Group\ Size_{g,t}$ - $5)$	2.75**			
	(1.21)			
$Sanctions_{h\neq g,t-1}$	8.05*	7.63*	8.22*	7.27**
	(4.15)	(3.65)	(4.17)	(3.73)
$Group\ Contribution_{h\neq g,t-1}$ if $Sanctions_{h\neq g,t-1} = 0$	-4.07*	-4.00*	-4.34*	-4.28**
	(2.26)	(2.12)	(2.31)	(2.14)
$Group\ Contribution_{h\neq g,t-1}$ if $Sanctions_{h\neq g,t-1} = 1$	2.05	2.06	2.06	2.11
	(2.07)	(2.10)	(2.14)	(2.05)
$\mathbb{1}(Group\ Size_{g,t} < 5)(5 - Group\ Size_{g,t})$		3.11*	3.20*	1.67
		(1.86)	(1.97)	(1.84)
$\mathbb{1}(Group\ Size_{g,t} > 5)(Group\ Size_{g,t} - 5)$		1.61	1.77	1.35
		(2.09)	(2.31)	(2.46)
Time Trend			3.98***	3.11*
			(1.54)	(1.68)
$Vote_{i,t-1}$				38.44***
				(5.41)
Observations	1,250	1,250	1,250	1,000

Notes. The table replicates the analyses in Table 5, reporting marginal effects of the covariates on the probability of voting for institutions (in percentage points) in the *Voting* treatment, using a control function approach for random effects panel probit models (Giles and Murtazashvili, 2013). Hausman tests justify the need to control for the endogeneity of $Initial\ Group_{i,t}$, $abs(Group\ Size_{g,t}$ - $5)$, $\mathbb{1}(Group\ Size_{g,t} < 5)(5$ - $Group\ Size_{g,t})$ and $\mathbb{1}(Group\ Size_{g,t} > 5)(Group\ Size_{g,t}$ - $5)$ (p-values<0.01 across all specifications). We used as instruments the fourth and fifth lags of the average contributions of other members of i's group (Smith, 2013), an indicator for subjects who moved before one of the five periods in which a voting stage takes place, and dummies for subjects who contributed more than 35 tokens in the first activity of the experiment - as a proxy for their "behavioral type" (Gürerk et al., 2006). Diagnostic tests indicate that the instruments are valid and not weak (the first-stage F-statistics are always higher than 10) (Stock and Yogo, 2005; Wooldridge, 2010). As required by Giles and Murtazashvili (2013)'s method, we fitted linear first-stage regressions for $Initial\ Group_{i,t}$, which work well in practice for dealing with binary endogenous covariates (Angrist, 2001). As an alternative, we used the generalized residuals of probit models as control variables (Fernández-Val and Vella, 2011), which are better suited for dealing with non-continuously distributed endogenous variables; the results are qualitatively similar. Bootstrapped standard errors adjusting for the first-stage estimation (Giles and Murtazashvili, 2013) are reported in parentheses. Significance levels: *** at 1%, ** at 5%, * at 10%. The first-stage estimates are available from the authors upon request.