TIPPING VERSUS COOPERATING TO SUPPLY A PUBLIC GOOD

Scott Barrett
Columbia University

Astrid Dannenberg
University of Kassel,
and University of Gothenburg

Abstract

In some important multiplayer situations, such as efforts to supply a global public good, players can choose the game they want to play. In this paper we conduct an experimental test of the decision by a group with fixed membership, playing over a finite number of periods, to choose between a "tipping" game, in which every player wants to contribute to the public good provided enough other players contribute, and a prisoners' dilemma, the classic cooperation game. In the prisoners' dilemma, the first best outcome is attainable, but cannot be sustained as a Nash equilibrium. In the tipping game, only a second best outcome may be attainable, but there exists a Nash equilibrium that is strictly preferred to the one in the prisoners' dilemma. We show that many groups persistently choose the prisoners' dilemma despite its strategic disadvantage, and that the groups that eventually choose the tipping game do better than the ones that stick with the prisoners' dilemma. (JEL: C72, C92, F53, H41)

1. Introduction

In "tipping" games, players behave differently either side of a "tipping point". In this paper, tipping represents a way of supplying a public good. On one side of the tipping point, no player wants to supply the public good; on the other side, every player wants to supply it. The tipping point thus represents a critical number of providers of the public good. Tipping is to be contrasted with the usual approach to supplying a public good, represented by the prisoners' dilemma, in which every player has a dominant

---

strategy not to contribute.[1] We report the results of an experiment in which members of a group vote to choose which of these games to play, knowing that a majority decides. The choice they face is difficult. The prisoners' dilemma can potentially achieve the overall first best outcome, but it cannot support this outcome as a Nash equilibrium. The tipping game may be able to support a Nash equilibrium that is Pareto-superior to the one in the prisoners' dilemma, but selection of this equilibrium may not be assured, and choice of this game might also foreclose any chance of attaining the first best outcome.

The prime example of the situation we have in mind is the provision of a *global* public good. Before countries choose how to play (choose their contributions to the public good) they must first agree on the rules of the game. For example, should they impose limits on the emissions of a pollutant or should they mandate a technology standard, the adoption of which would cause emissions to fall? The first approach is direct and, in the absence of any penalties for violations, leaves the prisoners' dilemma unchanged. The second approach, under the right conditions, is strategic and can turn the prisoners' dilemma into a tipping game. For example, if the adoption of a new technology entailed substantial network externalities, then it would pay every country to adopt the technology as soon as a critical group of other countries adopted it. Of course, countries may be reluctant to adopt the new technology if they fear that this critical group will never form. Moreover, technology standards are rarely the most cost-effective way to meet a particular environmental goal. Adoption of emission limits, by contrast, allows parties the flexibility to meet their obligations using the most cost-effective means. However, an agreement specifying emission limits leaves the prisoners' dilemma unchanged and so may have difficulties deterring free riding. Both approaches have been tried in the past to address a number of issues, ranging from climate change to ozone depletion to pollution of the seas (see Section 5). But which approach is best? This is a difficult question to answer in general because of the lack of a counterfactual: we do not know what would have happened had these issues been addressed differently. By studying behavior in the lab, however, we can observe the outcomes realized by groups that choose differently. We can also observe whether any groups reverse their decision at a later point in time and how these groups fare as a consequence.

As the tipping game has different strengths and weaknesses than the prisoners' dilemma, it may not be obvious which one will turn out to be the better choice in the end. Tipping games have multiple equilibria, and it may be difficult for the players to coordinate their behavior. As we explain later, coordination is especially difficult in the tipping game that is capable of sustaining only a second best outcome. The prisoners' dilemma, by contrast, has a unique equilibrium, but decades of experimental research have shown that many people do not play the equilibrium strategy, at least not at the

---

1. Tipping also differs from the much-studied threshold public goods game in which, once a threshold number of players has contributed, none of the other players wants to contribute to the public good (for a review of threshold public goods experiments, see Croson and Marks 2000).

beginning of the game (Ledyard 1995; Chaudhuri 2011). Forming expectations is thus difficult in both games, but giving the players the opportunity to update their beliefs may enable them to make better decisions over time.

Previous experiments on the endogenous choice of institutions have shown that individuals and groups tend to choose naïvely at first (for a review of this literature, see Dal Bó 2014). However, when given the opportunity to revise their initial choice, players often move gradually toward the welfare-improving institution. For instance, a number of studies have shown that players come to support the option to reward or punish other players and with this improve the prospects of cooperation (e.g., Gürerk, Irlenbusch, and Rockenbach 2006; Ertan, Page, and Putterman 2009; Sutter, Haigner, and Kocher 2010; Markussen, Putterman, and Tyran 2014).[2]

Our experiment comes closest to two recent experiments conducted by Dal Bó, Foster, and Putterman (2010) and Dal Bó, Dal Bó, and Eyster (2013). In both of these experiments, players can vote to modify the payoffs of a two-person prisoners' dilemma. In Dal Bó et al. (2010), the players can vote for a fine to be imposed on unilateral defection, an off-equilibrium change that makes mutual cooperation another Nash equilibrium of the game while leaving the payoffs to mutual cooperation unchanged. In this experiment, groups that voted for the change earned higher payoffs, but only about half the groups voted this way. However, because the players were not allowed to revise their choice, we do not know if they would have chosen differently in a second voting round.[3] In Dal Bó et al. (2013), the players can vote for a fine that reduces the payoff to playing *every* strategy, but with the payoff to defection falling by more than the payoff to cooperation. In this alternative game, mutual cooperation has a lower payoff than in the original two-player prisoners' dilemma but cooperation becomes the dominant strategy for both players. As in the previous paper, groups that voted for the change earned higher payoffs, but only about half the subjects voted this way. However, in a treatment that allowed subjects to vote repeatedly before each of the five rounds of play (Majority Repeated), the players learned to overcome their bias in favor of the prisoners' dilemma. By the end of this treatment, only two out of twenty groups were still playing the prisoners' dilemma.

In our experiment, choice of a regime is harder than in Dal Bó et al. (2010) because going for the tipping regime may mean foregoing the opportunity to realize a higher payoff in the prisoners' dilemma. Choice of a regime in our experiment is also harder than in Dal Bó et al. (2013) because our alternative game has two Nash equilibria, only one of which offers a higher payoff compared to the Nash equilibrium in the prisoners' dilemma. Unlike both of these papers, we compare treatments representing different alternatives to the prisoners' dilemma. One of these alternatives is capable of

---

2.   Our main concern in this paper is with games played by nation states. Focusing on cases in which states have used sanctions, Hufbauer et al. (2007, p. 7) observe that, for many reasons, "Sanctions often do not succeed in changing the behavior of foreign countries."

3.   The primary aim of Dal Bó et al. (2010) is to show that a regime imposing the fine has a bigger effect on behavior when it is chosen by the players who will ultimately be subject to the fine than when it is imposed upon these players without their consent.

sustaining the overall first best outcome and aligns payoff- and risk-dominance. The other one is capable of sustaining only a second best outcome and creates a conflict between payoff- and risk-dominance. By comparing these treatments, we are able to show whether and how behavior differs depending on the nature of the alternative. Moreover, in our experiment, five players vote for which game to play and then play the chosen game *as a group*—a context that is particularly suited to understanding negotiations of multilateral international agreements. In the experiments noted above, by contrast, a group of four (in Dal Bó et al. 2010) or six (in Dal Bó et al. 2013) players chooses which game to play, with *pairs* of players then playing the chosen game—a context that is more suited to the study of domestic law making. Finally, we also make public the results of each vote, a design feature that should help the players to coordinate. This assumption is consistent with the way multilateral negotiations are conducted,[4] but would of course be inappropriate for the study of a "democracy" in which the final vote tally is public knowledge but the voting decisions of particular individuals are private.

We find that every group that chooses to play the tipping game is able to coordinate perfectly, sustaining a 100% group contribution level. By contrast, the groups that choose to play the prisoners' dilemma cooperate only partially and at a decreasing rate. As a consequence, the groups that choose the tipping game do better than the groups that choose the prisoners' dilemma, even when the tipping game can sustain only a second best outcome. Our key result, however, is that whether players *choose* to play the tipping game depends crucially on the nature of this game. Similar to Dal Bò et al. (2013) and other experiments on endogenous institutions, we find that, when the tipping game can sustain the first best outcome and the mutually preferred equilibrium is both payoff- and risk-dominant, players are initially unsure of which game to play but, over time, they all move toward the regime that pays off more handsomely. By contrast, when the tipping game can sustain only a second best outcome, and the payoff-inferior equilibrium is risk-dominant, only half the groups move to this regime. The other groups remain in the prisoners' dilemma throughout the entire experiment even though they are unable to sustain much cooperation. This degree of "stickiness" has not been observed previously.

Much of our paper is devoted to revealing the *reasons* different groups behave differently. We show that the groups that switch to the tipping game and the groups that never switch have different expectations. In order to switch, players must come to believe two things—first, that cooperation in the prisoners' dilemma will fail and, second, that coordination on the mutually preferred equilibrium in the tipping game will succeed. Low expectations for the tipping game explains the puzzling result that groups remain stuck in the prisoners' dilemma even though they cannot sustain much cooperation.

---

4. For example, the rules of procedure for the Montreal Protocol say that substantive decisions are to be made by a show of hands or a rollcall vote (see http://ozone.unep.org/Publications/ VC_Handbook/Section_3_Rules_of_Procedure/Rules_of_procedure.shtml. Accessed on 30 January, 2017).

We also find that some people are better than others at reading the signals conveyed by the way people vote. Voting to play the tipping game signals an intention to contribute to the public good provided that the tipping game is chosen. Some players read these signals very well; their behavior appears strategic. Other players miss the signals, but if made to play the tipping game these players come to see that once a majority of co-players contribute it is in their best interest to contribute, too. These players' behavior appears adaptive.

We cannot be sure that the groups that remain in the prisoners' dilemma would do better by switching. After all, these groups' low expectations for the tipping game could become a self-fulfilling prophecy. However, and as noted above, we find that individuals who vote to remain in the prisoners' dilemma, but are out-voted by their group members, learn to coordinate on the better equilibrium in the tipping game. As a consequence, these individuals end up getting a higher payoff than the players who vote the same way (i.e., for prisoners' dilemma) but, being in a majority, get to play their preferred game. To explore the role of expectations in more depth, we developed a third treatment in which groups must play both games before being given the opportunity to choose which game to play. This earlier experience, we find, causes the players to update their expectations, leading more groups to choose the tipping game compared to the treatment in which groups lack this experience. Taken together, our findings suggest that if the expectations of groups can be shifted to create a majority in favor of the tipping game, then this group will not only choose to play the tipping game but, having made this choice, will also coordinate successfully and, as a consequence, earn a higher payoff.

In the next two sections we present our underlying model and describe our experimental design and main treatments. In Section 4 we present our main results on the choices made by individuals and groups, and show how these choices are shaped by expectations and learning. In Section 5 we use our results to interpret several real world examples of agreements to supply a global public good. We conclude with some final observations on our main results.

2. Model

Our experiment is based on a particular representation of a more general model. In the more general model, there are N players, all of whom have identical payoff functions. In the metagame, the players first choose which game to play, the A game or the B game. They then play the game they have chosen. We begin by describing these individual games.

In the A game, the players have a binary choice; every player i ($i = 1,..., N$) must choose $q_i \in \{0, 1\}$, taking as given the choices made by the other players. Letting k denote the number of *other* players that play $q_j = 1$, i's payoff is assumed to be given by

$$\pi_i^A (1; k) = b (k + 1) - c, \pi_i^A (0; k) = bk, \tag{1}$$

with $bN > c > b > 0$. In this game, play $q_i^* = 0$ is the unique Nash equilibrium, but full cooperation requires that every player i plays $q_i^{FC} = 1$. This is a prisoners' dilemma.

In the B game, the players have another binary choice; every player i must choose $y_i \in \{0, 1\}$. Letting m denote the number of *other* players that choose $y_j = 1$, player i's payoff is assumed to be given by

$$\pi_i^B (1; m) = b (m + 1) - c - d, \, \pi_i^B (0; m) = \alpha m. \qquad (2)$$

The parameter d represents the cost-penalty to playing the tipping game as compared to the prisoners' dilemma. Assume $d \geq 0$, $N > (c + d - \alpha)/(b - \alpha)$, and $c + d > b > \alpha \geq 0$. It is then easy to show that $\pi_i^B(1; m) > \pi_i^B(0; m)$ for $m > \tau$ and $\pi_i^B(1; m) < \pi_i^B(0; m)$ for $m < \tau$, where $\tau = (c + d - b)/(b - \alpha)$. τ thus represents the "tipping point" for the B game. Our assumptions about the parameters imply $\tau \in (0, N)$.

In the B game, there are two Nash equilibria in pure strategies. In one, every player plays $y_i^* = 0$. In the other, every player plays $y_i^{**} = 1$. All players earn strictly higher payoffs in this second pure-strategy Nash equilibrium compared to the first one.[5] However, and as explained more fully in the next section, it is not obvious that the players will be able to coordinate on this second equilibrium. Coordination on the payoff-dominant equilibrium may depend on whether this equilibrium is also risk-dominant. More generally, equilibrium selection will depend on expectations. In our experiment, the games are played a finite number of times, allowing expectations to be updated depending on which game was played in the past and how it was played.

Since each player gets a payoff of zero in the Nash equilibrium of the A game and a payoff of zero or greater (in particular, either zero or $bN - c - d > 0$) in the B game, depending on which (pure strategy) Nash equilibrium is selected, it might seem that rational players should choose B over A. The reason people may not choose this way is that people often cooperate in the prisoners' dilemma, at least in the beginning of the game (Ledyard 1995). The Nash equilibrium assumes that the players are rational and have self-regarding preferences, and that this is common knowledge. However, many people seem willing to contribute if they expect that their co-players will contribute or if they observe that their co-players have contributed in the past—a possibility that may also be common knowledge. Because coordination on the efficient equilibrium in the B game may fail and because cooperation in the A game may succeed, at least partially, especially in the early contribution rounds, it is not obvious that the players in our experiment will *choose* B over A, or that they would *do better* by choosing B over A in the metagame. Moreover, because expectations affect how these games

---

5. There also exists a mixed strategy equilibrium in which every player earns an expected payoff somewhere in between the payoffs corresponding to these pure strategy equilibria. Letting p denote the probability, from every player i's perspective, that each player j, $j \neq i$, will play $y_j = 1$, the mixed strategy equilibrium involves each player choosing to contribute with probability $p^* = \tau/(N - 1)$, yielding each player an expected payoff $E(\pi_i^B) = \alpha \tau$. It is easy to confirm that $\pi_i^B(1; N - 1) > E(\pi_i^B(p^*; p^*(N - 1))) \geq \pi_i^B(0; 0)$.
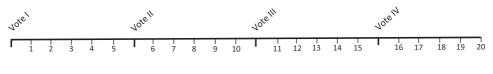
FIGURE 1. Vote-First Treatments.

are played, and because expectations may change over time, depending on how these games were played in the past, the history of play may affect how the players choose between A and B in the metagame. Our experiment is designed to investigate all of these issues.

3. Experimental Design

Our experiment is played by groups of $N = 5$ players. In the metagame, each group must choose between playing the A game, a prisoners' dilemma, and the B game, a tipping game. The A game is the same in all of our treatments. The B game, however, varies with the treatment. In the treatment *Vote-First-B-10*, groups choose between A and B-10. In *Vote-First-B-8*, they choose between A and B-8. One difference between these treatments is that the Pareto-superior (pure strategy) Nash equilibrium in B-10 yields the same payoff as the full cooperative outcome in the A game, whereas the Pareto-superior Nash equilibrium in B-8 yields a lower payoff compared to the full cooperative outcome in the A game. Another difference is that the payoff-dominant Nash equilibrium is risk-dominant in B-10 and not risk-dominant in B-8.

In both treatments, the experiment is played in four phases; see Figure 1. At the start of each phase, the players vote to choose the game they want to play, with a simple majority deciding.[6] Afterward, they play the chosen game in five consecutive contribution rounds, with all the players choosing (simultaneously) whether to contribute to the public good in each round. Since there are four phases, there are 20 contribution rounds in total. It is common knowledge that individual votes are made public to all the players after each voting round and that individual contribution decisions are made public after each contribution round.

In every contribution round, each player is given two playing cards, one red and one black, and must decide which card to return. If player i is playing the A game, returning the red (black) card is equivalent to choosing $q_i = 1$ ($q_i = 0$). If playing the B game, returning the red (black) card is equivalent to choosing $y_i = 1$ ($y_i = 0$). In both cases, handing back the red card supplies the public good.

Every player's payoff, relative to the theoretical model, is increased by an amount s. This scaling has no effect on the theory, but is needed to ensure that players cannot be left out of pocket when playing the experiment. In the A game, players get $s - c$ if

---

6. In international negotiations, decisions of this kind are routinely made by a vote. For example, the rules of procedure for meetings of the parties to the ozone agreements say that "decisions . . . on all matters of substance shall be taken by a two-thirds majority vote. . . ." (see http://ozone.unep.org/ Publications/VC_Handbook/Section_3_Rules_of_Procedure/Rules_of_procedure.shtml).
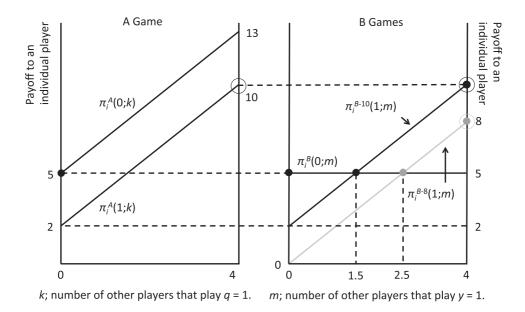
FIGURE 2. The A and B games.

they hand in their red card and s if they hand in their black card. Either way, they get b for every red card handed in by anyone in the group.

In both versions of the B game, players who hand in their black card get a payoff of s plus an amount α for every red card handed in, whereas players who hand in their red card get a payoff $s - c - d$ plus an amount b for every red card handed in. The difference between B-10 and B-8 is reflected in the value of d. Our experiments assume $d = 0$ for B-10, and $d = 2$ for B-8. All other parameter values are the same values in both treatments: $\alpha = 0$, $b = 2$, $c = 5$, and $s = 5$. Obviously, different parameter values could be chosen; in particular, α could be positive or negative.[7] However, this particular combination of values is simple and ensures that the tipping points for the two B games are equidistant from, but on opposite sides of, $(N - 1)/2$. [8]

The A and B (that is, B-10 and B-8) games are shown in Figure 2.[9] Here it can be seen that the "10" in B-10 and the "8" in B-8 represent, respectively, the full

---

7. Given our assumptions, $\alpha > 0$ yields a tipping game provided $bN - c - d > \alpha(N - 1)$; $\alpha \leq 0$ always yields a tipping game. A positive value for α would be consistent with a weak network externality in the choice of $y_i = 0$ over $y_i = 1$. A negative value for α would be consistent with a very strong network externality. Obviously, our model assumes that payoffs are linear, but they could be nonlinear. For a discussion of the possibilities, see Schelling (1978).

8. Setting $c = 5$ and $b = 2$, we have $\tau = (3 + d)/(2 - \alpha)$. Letting $\alpha = 0$ and recalling that $d = 0$ for B-10, we get $\tau = 3/2$; for B-8, $d = 2$, giving $\tau = 5/2$. If the players reason that others are as likely to contribute as not to contribute, they will evaluate the payoff to contributing or not at $(N - 1)/2$, which for $N = 5$ yields $4/2$. Obviously, the tipping points for B-8 and B-10 are, respectively, 1/2 above and below this value.

9. This kind of figure was first developed by Schelling (1978).

A	Number of red cards handed in by your co-players				
	0	1	2	3	4
Red	2	4	6	8	10
Black	5	7	9	11	13

B-10	Number of red cards handed in by your co-players				
	0	1	2	3	4
Red	2	4	6	8	10
Black	5	5	5	5	5

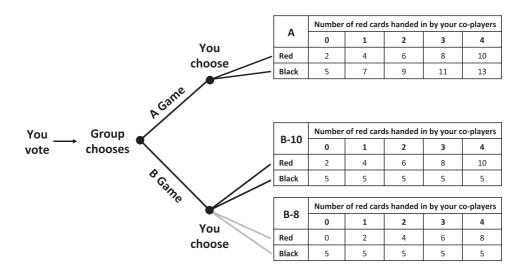| B-8	Number of red cards handed in by your co-players				
	0	1	2	3	4
Red	0	2	4	6	8
Black	5	5	5	5	5

FIGURE 3. Voting stage and payoffs.

cooperative payoffs in these two games (the full cooperative payoff in the A game is 10). Note as well that the closed dots in Figure 2 represent Nash equilibria (the mixed strategy equilibria of the B games are "interior"), and the open circles represent the efficient outcomes for the different games. The payoffs are shown in Figure 3.

For which game will people cast their vote? As noted before, the Pareto-inefficient Nash equilibria in the B games are no worse than the unique Nash equilibrium in the A game, whereas the Pareto-efficient pure strategy equilibrium in both B games is strictly preferred by all players to the Nash equilibrium of the A game. It might thus seem that the players should vote for B. However, closer inspection reveals a more complex picture.

Some players might form a "first impression" of which game they should choose by looking at the payoffs. From this perspective, there are many reasons for players to prefer A to B. In the *Vote-First-B-10* treatment, for example, the lowest individual payoff is the same in the A and B games, whereas A pays out the highest individual payoff. Similarly, in *Vote-First B-8*, the lowest and the highest individual payoffs are both strictly higher in the A game than in the B game.[10]

Other players might look more deeply into these games, trying to reason through how their co-players will play. For example, in B-10, they might see that coordination on the welfare superior equilibrium in the B game seems likely given that contributing

---

10.   Note also that in *Vote-First-B-10*, for any given contribution decisions, the aggregate payoff is never lower and is often higher in the A game. Also, a person who intends not to contribute does at least as well opting for A, whereas a person who intends to contribute is indifferent between A and B. Similarly, in *Vote-First-B-8*, players might be drawn to A because it offers the highest collective payoff. They might also notice that a person intending to contribute is strictly worse off when playing the B game than the A game, given the choices by the other players of whether or not to contribute, and that a person intending not to contribute does at least as well choosing A as B.

in this game is both payoff-dominant and risk-dominant (if each player believes that the other players are as likely to contribute as not, then each player can expect that two other players will contribute, in which case each player can expect to get a payoff of 6 by contributing and a payoff of 5 by not contributing). In B-8, reasoning through how others will play is more difficult. The tipping point is higher for B-8 than for B-10. Moreover, the Pareto inefficient Nash equilibrium is risk-dominant. For both reasons, coordination on the Pareto-efficient equilibrium seems less sure in this game.[11]

However, and as noted before, in our experiment individual votes are public knowledge. Votes not only determine the game that is chosen; they also serve as a signal for subsequent contribution decisions. This signaling should be particularly useful in the B games, where the simple majority (at least three out of five) is equal to or greater than the tipping point. In both treatments, it makes the most sense for players to vote for B if they think coordination on the mutually preferred equilibrium will succeed. But the players who believe coordination will succeed should then contribute when B is chosen. Hence, *all* the B-voters should contribute when B is chosen. But then the A-voters should expect that all the B-voters will contribute, making it in *their* interest to contribute, too. In other words, with vote signaling, players should expect that coordination on the efficient equilibrium will succeed, even for the B-8 treatment. However, this reasoning demands an unusual degree of sophistication. Some players might reason through their decision problem in this way, but others might go with their "first impression".

4. Experimental Results

The experimental sessions were held in a computer lab at the University of Magdeburg, Germany, using undergraduate students recruited from the general student population. In total, 300 students participated in the experiment, each student taking part in one treatment only. There were three treatments (the two *Vote-First* treatments discussed previously and one *Play-First* treatment discussed in Section 4.5) with 20 groups per treatment and five players per group.

The experimental instructions handed out to the students included several numerical examples and control questions.[12] The control questions tested subjects' understanding of the game to ensure that they were aware of the available strategies and the implications of making different choices. After reading the instructions and answering the control questions correctly, subjects began playing the game. In each session, 25

---

11. The central problem here is "strategic uncertainty", meaning that each player is uncertain about the equilibrium strategy that will be chosen by the other players. Different theoretical approaches have been applied to this problem, including traditional noncooperative game theory, evolutionary game theory (which emphasizes dynamics and learning), and adaptive learning (in which players' strategic sophistication is assumed to be limited). See Crawford (1997) for a review of this literature and its relationship to experimental work. For a review of experiments in coordination games, including discussion of conflicting evidence for equilibrium selection based on payoff- and risk-dominance, see Devetag and Ortmann (2007).

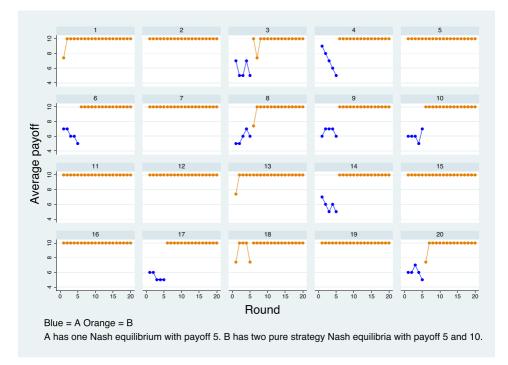12. The experimental instructions are provided in the Appendix.

Blue = A Orange = B

A has one Nash equilibrium with payoff 5. B has two pure strategy Nash equilibria with payoff 5 and 10.

FIGURE 4. Payoffs over time by group for *Vote-First-B-10*.

subjects were seated at linked computers (game software *z*-Ttree; Fischbacher 2007) and randomly assigned to one of 5 five-person groups. The subjects did not know the identities of their co-players, but they did know that the membership of their group remained unchanged throughout the session. To ensure anonymity, each individual within a group was identified by a different number, from 1 to 5. During the game, earnings were displayed in tokens. It was public knowledge that payments would be calculated by summing up the number of tokens earned over all 20 contribution rounds and by then applying an exchange rate of €0.10 per token. Before and after the game, the subjects were asked to complete questionnaires.[13] After the final questionnaire was completed, the subjects were paid their earnings privately in cash.

Our main results for the *Vote-First* treatments are shown in Figures 4 and 5 and summarized in Table 1. Figures 4 and 5 show the average payoff per contribution round for each group (of which there are 20 in total), depending on the game chosen by the group (A in blue, B in orange) over the four phases.[14] A quick look at the figures shows that behavior differs dramatically between *Vote-First-B-10* and *Vote-First-B-8*. We discuss these differences in detail below.

---

13.   The postplay questionnaire results are discussed in Section 4.4; the preplay questionnaire results are discussed in footnote 24.

14.   We only show payoffs as the figures for contributions reveal a nearly identical pattern.

Blue = A Orange = B
A has one Nash equilibrium with payoff 5. B has two pure strategy Nash equilibria with payoff 5 and 8.

FIGURE 5. Payoffs over time by group for *Vote-First-B-8*.

TABLE 1. *Vote-First* treatments.

		Vote-First-B-10			*Vote-First-B-8*		
Phase	Game	% of groups	Average % contributions	Average payoff	% of groups	Average % contributions	Average payoff
I	A	45	21	6.1	100	39	7.0
	B	55	99	9.8	0	–	–
II	A	0	–	–	90	24	6.2
	B	100	100	9.9	10	90	7.2
III	A	0	–	–	80	26	6.3
	B	100	100	10	20	90	7.1
IV	A	0	–	–	55	10	5.5
	B	100	100	10	45	94	7.5

4.1. *Voting*

Consider, to begin, the voting behavior of individuals, summarized in Figure 6. In *Vote-First-B-10*, 57% of players voted for game B in the initial phase, rising to 91%

FIGURE 6. Individual voting decisions in the *Vote-First* treatments.

by the fourth and final phase.[15] In *Vote-First-B-8*, 11% of the players voted for B initially, rising to 51% by the final phase. The switching behavior of all individuals taken together (in favor of B) is similar for the two treatments, but the initial support given to A rather than B differs greatly.

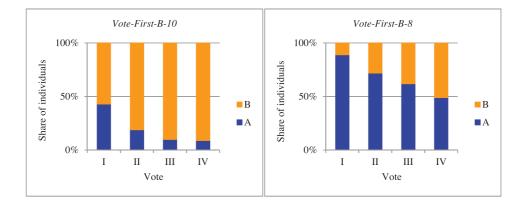The behavior of individuals is consistent with these aggregate observations. In *Vote-First-B-10*, 36% of the players started by voting for A, and then switched to B at some point without ever switching back, whereas in *Vote-First-B-8*, 37% of the players voted this way. In *Vote-First-B-10*, 4% switched from B to A before switching back to B, whereas in *Vote-First-B-8*, 7% did this. Finally, in *Vote-First-B-10*, 6% switched from B to A without ever switching back, whereas in *Vote-First-B-8*, 11% behaved in this same way. Again, the main difference in behavior is reflected in the "core" support for A rather than B. In *Vote-First-B-10*, 51% of the players voted for B every time, whereas just 3% voted for A every time. In *Vote-First-B-8*, this behavior was almost reversed, with 38% voting for A every time and just 7% voting for B every time. To summarize:

Result 1. *In Vote-First-B-10 there is core support for B, whereas in Vote-First-B-8 there is core support for A. Vote switching behavior is very similar for the two treatments, with the vast majority of switchers moving from A to B.*

This individual voting behavior is reflected in the choices made at the group level. In *Vote-First-B-10*, 11 out of 20 groups started by playing B and never switched. The other nine groups initially gave their support to A, but all of these groups switched to B at the next opportunity and never switched back. In *Vote-First-B-8*, *all* groups started by playing A. In the second phase, two groups switched to B, but then subsequently

---

15.   Individual voting behavior in the initial phase of *Vote-First-B-10* is surprisingly similar to the results observed by Dal Bó, Foster, and Putterman (2010). In their experiment, 54% of players voted to play the coordination game. However, as we show here, support for the coordination game quickly increases when players have the chance to revise their choice.

switched back to A before returning to B again in the last phase. Four groups switched to B in the third phase and another four groups persisted in playing A until the last phase, when they finally switched to B. Ten groups *never* chose B. These groups were strongly attracted to A and/or repelled by B (we discuss these effects later). Overall, the difference in group-behavior between the two treatments is highly significant. The proportion of groups choosing to play B is significantly higher in *Vote-First-B-10* than in *Vote-First-B-8* (Fisher's exact test, $p < 0.01$ for each phase).[16]

Result 2. *In Vote-First-B-10, groups were initially divided in their support for A and B, but support quickly shifted to B; ultimately, group support for B was universal. In Vote-First-B-8, all groups started out supporting A, but over time about half the groups switched to B; the other groups never chose B.*

4.2. Contributions and Payoffs

In *Vote-First-B-10*, the groups that chose A in the first phase cooperated only partially; 31% of the subjects contributed in the first contribution round, declining to 9% by the fifth round, for an average of 21% (see Table 1). The groups that chose B in the first phase of *Vote-First-B-10*, by contrast, started out making high contributions and then increased these, quickly reaching the maximum level. Taking the group average for the first phase as the unit of observation, a Mann–Whitney–Wilcoxon (MWW) test shows that the difference in contributions between the groups that played A and the groups that played B is highly significant ($p < 0.01$). The players in *Vote-First-B-10* also received a higher average payoff when they played B than when they played A ($p < 0.01$).

Over all four phases of *Vote-First-B-8*, contributions in the A game (averaged over all groups playing A) generally declined (see Table 1).[17] Contributions in the first phase of the A game averaged 39%, dropping to 10% by the last phase. Contributions started at 62% in the first contribution round, declining to 5% by the 20th round. As in *Vote-First-B-10*, contributions in the B-8 game settled at the optimal level by the end of every phase in which B was played. Also, following each vote, average contributions for the groups playing B are always significantly higher than for the groups playing A (MWW test, $p < 0.05$ for each phase). Average payoffs are also higher for the groups that chose B rather than A. Here, the differences are weakly significant for the second and third phases ($p < 0.10$), and highly significant for the last phase ($p < 0.01$).

Result 3. *For both of the Vote-First treatments, contributions and payoffs are significantly higher when groups play B than when they play A.*

The behavior of Group 25 (see Figure 5) demonstrates the allure of the A game in *Vote-First-B-8*. The players in this group are initially drawn to A, perhaps because

---

16.   Unless stated otherwise, all statistical tests reported in this paper are two-sided and take the group as the unit of observation—a conservative approach.

17.   As shown in Figure 5, contributions by individual groups reflect a similar pattern.

playing A has the potential of yielding a higher payoff or perhaps because they are unsure as to whether coordination in the B game will succeed. The problem is that the potential to earn more when playing A can only be realized if all of the players contribute when playing A, and the group is unable to sustain much cooperation for long. By contrast, these same players coordinate flawlessly when the group switches to B after the second voting round. In the third voting round, the group switches back to A, presumably in the belief that contributions will remain high after the switch. However, contributions collapse in much the same way as before, and in the final voting round the group switches back to B, settling for the second best. Many other groups behave similarly. However, about half do not—they persist in playing A, even though they also cannot sustain much cooperation.

Together, Results 2 and 3 imply:

Result 4. *In Vote-First-B-10, all groups converge quickly to the B game, and then coordinate flawlessly, sustaining the first best outcome. In Vote-First-B-8, about half the groups move to the B game, eventually coordinating flawlessly and sustaining the second best outcome. The other groups remain in the A game even though they sustain little cooperation.*

4.3. Expectations

What are the reasons some groups remain in A and some switch to B in *Vote-First-B-8*? The analysis developed here draws from varying kinds of evidence, and yields a strong result:

Result 5. *In Vote-First-B-8, whether groups switch to B or persist in playing A depends on two different forces. Low expectations for successful cooperation "push" groups away from A. High expectations for successful coordination "pull" groups toward B. Both effects are necessary; neither is sufficient.*

Our evidence for this result is of three types. First, we are able to infer expectations from the choices observed in the games. This evidence is provided below. Second, we asked the players in an ex post questionnaire what they expected and what motivated their choices. This evidence is provided in Section 4.4. Finally, we conducted an additional treatment, called *Play-First*, in which we manipulated expectations by requiring that groups gain experience playing both games before choosing the game they would prefer to play. This evidence is provided in Section 4.5.

Regarding the first type of evidence, note first that, in *Vote-First-B-8*, *all* groups initially chose to play A, a group behavior that reflects an expectation by a majority that payoffs will be higher when playing A than when playing B. Plainly, groups must have a disappointing experience playing A before being willing to try B. The push effect is thus necessary for getting players to move to B.[18]

---

18.   There is, of course, a chance that had individuals been sorted differently, at least one group would have comprised a majority of first-time B voters. In our experiment, out of 100 players in the *Vote-First-B-8*

TABLE 2. Probit regression on voting for and selecting A in *Vote-First-B-8*.

	Individual level		Group level
Variables	(1) Voting decision (A = 1, B = 0)	(2) Voting decision (A = 1, B = 0)	(3) Game selection (A = 1, B = 0)
Lagged voting decision	1.339*** (0.221)	1.351*** (0.219)	
Lagged individual payoff in A	0.228*** (0.068)		
Lagged group contribution in A		1.645** (0.826)	6.966** (3.657)
Constant	−2.535*** (0.610)	−1.603*** (0.504)	−0.843 (0.849)
Observations	270	270	54
Number of subjects	100	100	
Number of groups			20

Notes: Random effects probit regression. Standard errors in parentheses. Significance levels: ***$p < 0.01$; **$p < 0.05$. Regressions at the individual level include group dummies, which are insignificant and not shown here. Dependent variables: voting decision = 1 if player voted for A in current phase, 0 otherwise. Game selection = 1 if group played A in current phase, 0 otherwise. Independent variables: lagged voting decision = 1 if player voted for A in previous phase, 0 otherwise. Lagged individual payoff in A = player's average payoff in the previous phase of playing A. Lagged group contribution in A = group's average contribution in previous phase of playing A.

We can also show that a stronger push effect increases the likelihood of any group moving to B. Table 2 presents results from three probit regressions. Columns (1) and (2) show regressions of the decision by individuals to vote for A in any phase, conditional on this individual having played A in the previous phase.[19] The results reveal voting inertia: individuals tend to vote for A if they voted for A previously. This was to be expected since, as noted previously, once individuals vote for B they rarely switch to A. However, the results also reveal that the decision to vote for A depends strongly on the experience associated with having played A previously. The players that switched their vote to B had a particularly bad experience when playing A in the previous phase. Note that this effect is observed irrespective of whether an individual's experience is measured by his or her payoff when playing A (column (1)) or his or

treatment, 86 voted for A at the first opportunity and 14 voted for B. The probability that a group of five players drawn at random from this pool of 100 players will contain at least three first-time B-voters can be shown to be less than 2%.

19.   We excluded from this regression all the first-phase votes, which depend only on expectations. This leaves 300 observations (three phases times 100 players voting in each phase). We also excluded the 30 observations corresponding to situations in which groups played B in the previous phase, leaving 270 observations. We excluded these observations because the purpose of this regression is to examine voting behavior when subjects played the same game in the previous phase. However, if we include the groups that played B in the previous phase and take the outcome of the last phase in which these groups played A as an explanatory variable, the results are almost the same.

TABLE 3. Probit regression of individual contribution decisions in *Vote-First-B-8*.

Variables	Individual contribution decision (red = 1, black = 0)	
	(1) Game A in current phase	(2) Game B in current phase
Voting decision (A = 1, B = 0)	−0.359	−1.684***
	(0.215)	(0.476)
Lagged individual contribution in A	1.628***	0.0199
	(0.364)	(0.972)
Constant	−0.601	1.326**
	(0.403)	(0.614)
Observations	210	60
Number of subjects	90	50

Notes: Random effects probit regression. Standard errors in parentheses. Significance levels: ***$p < 0.01$; **$p < 0.05$. Regressions include group dummies, which are insignificant and not shown here. Dependent variable: contribution decision = 1 if player played red, 0 otherwise (only contribution decisions in the first round of the current phase of playing A or B are included). Independent variables: voting decision = 1 if player voted for A, 0 otherwise. Lagged individual contribution in A: average number of red cards contributed in the previous phase of playing A.

her group's average contribution level when playing A (column (2)). Finally, column (3) shows that this result also holds at the group level: groups are more likely to stick with A if they experienced a higher contribution level when playing A in the previous phase.

To summarize, in *Vote-First-B-8*, all groups are initially drawn to A. Some are later "pushed" into trying B if and when their experience in playing A proves particularly disappointing. This effect of getting groups to try B is crucial. Of the ten groups that tried B at some point, all but one ended up choosing B in the last voting round. Of the 11 groups that chose A in the last round, only one had ever tried playing B before.

We now turn to the pull effect. Recall from Section 3 that it can only pay players to vote for B if they believe coordination will succeed. But if players believe that coordination will succeed, then they will want to contribute when playing B. Hence, players who expect coordination to succeed when playing B should be more inclined to vote for B and to contribute when B is chosen. Players who expect coordination to fail should be less inclined to vote for B and to contribute when B is chosen. Table 3 presents a probit regression of individual contribution decisions in the first round of playing A (column (1)) or B (column (2)), conditional on this individual having played A in the previous phase.[20] The results reveal remarkable differences between the groups that play A and those that play B. For the groups that play A there is no significant

20.   Only in this first contribution round are expectations determined exclusively by the voting outcome and previous experience playing A. Again, we exclude from this regression all the first-phase observations as well as those corresponding to when B was played in the previous phase. In total, groups chose A after having played A in the preceding phase 42 times, making (since there are five players per group) 210

TABLE 4. Responses to the ex post questionnaire (% of subjects).

| | | Vote-First-B-10 | | Vote-First-B-8 | |
| | | Played B every time ($n = 55$) | Played A at least once ($n = 45$) | Played A every time ($n = 50$) | Played B at least once ($n = 50$) |
Question	Answer				
Did you expect	Very much	7	20	30	26
your fellow co-players	Somewhat	15	47	34	50
to hand in their	Little	47	18	32	18
red card in game A?	Not at all	30	16	4	6
Did you expect	Very much	98	96	38	84
your fellow co-players	Somewhat	2	4	30	4
to hand in their	Little	0	0	12	4
red card in game B?	Not at all	0	0	20	8
If you could give advice	Game A	4	0	82	26
to a new group of	Game B	96	100	18	74
participants, which game					
would you recommend					
that they play?					

difference between A-voters and B-voters. What drives their contribution decision is their decision when playing A previously: the lower a player's average contribution in the previous phase the less likely the player is to contribute in the first round of the next phase. In other words, free riders tend to remain free riders and cooperators tend to remain cooperators.

Interestingly, lagged contributions in the A game do not have a significant effect on contributions in the first round after a group has switched to B. Thus, whether a subject was a free rider or a cooperator in the A game did not affect how this player behaved in the B game. However, we find a significant difference in the contributions of the players who *vote* for A and the players who *vote* for B: B-voters are more likely to contribute than A-voters when playing the B game for the first time. This implies that B-voters must be more optimistic about coordination succeeding in the B game—presumably the reason they voted for B in the first place. This is the pull effect.

4.4. Ex Post Questionnaire

Table 4 presents responses by the players to a questionnaire given after they had finished playing.[21] In *Vote-First-B-10*, we distinguish between groups that played B every time and those that played A at least once (of course, in this treatment, no group played

---

observations; groups played B after having played A in the previous phase a total of 12 times, giving 60 observations.

21. Responses to an ex post questionnaire are likely to reflect both expectations and experience. However, a preplay questionnaire might have biased subsequent behavior in the game. Also, much of the dynamics occurred during the game and would not have been captured by a preplay questionnaire.

A more than once). In *Vote-First-B-8*, we distinguish between groups that played A every time and those that played B at least once.

Two observations stand out. First, expectations for successful coordination are very high in *Vote-First-B-10*. They are also high in *Vote-First-B-8* for the groups that played B at least once. Though we cannot exclude the possibility that these beliefs were formed while the game was played rather than before it was played, the beliefs expressed are at least consistent with the pull effect. By contrast, expectations for successful coordination are noticeably lower for the individuals in groups that never played B in *Vote-First-B-8*. Interestingly, these players' expectations for contribution levels overall are similar for the B game and the A game (compare their responses to the first two questions in the table).[22] All other players have very different expectations for the two games (for the first two questions, compare the responses of the players who played A every time in *Vote-First-B-8* with the responses of the other players).[23] It thus seems that the individuals who repeatedly voted for A did so in the belief that the outcome in B would be no better and might be worse—a belief that formed despite the strategic advantage of the B game and that did not change despite low cooperation in the A game.

Second, almost all the players in *Vote-First-B-10* would recommend that a new group of participants play B rather than A. By contrast, individuals who took part in *Vote-First-B-8* were divided. A large majority of those who played A every time would recommend A, whereas most of the players who played B at least once would recommend B—further confirmation of the pull effect.

We also asked our participants in an open-ended question to give the reason for their recommendation. Many of the players in *Vote-First-B-8* who played A every time and who also recommended that others play A said that, in their view, game A was the better game. A typical answer was, "I would recommend game A and wish them a more cooperative group than the one I had." These players seemed to believe that the level of cooperation was determined by the group (perhaps reflecting a belief that another group would comprise more players with other-regarding preferences) and not by the game. One of the players who played B in *Vote-First-B-8* and who also recommended B said this: "Play A with people you know and trust, but play B with people you don't know." This answer reflects a better strategic understanding of the different incentives created by the two games.[24]

---

22. Dal Bó et al. (2013) obtain a similar result. In their experiment, players who voted for the prisoners' dilemma were less likely to believe that behavior would be different for the two games.

23. A Wilcoxon signed-rank test shows that the difference in expectations for the A and B games is not statistically significant for the individuals in groups that never played B in *Vote-First-B-8* ($p = 0.98$) but that it is highly significant for the individuals in all other groups ($p < 0.01$ each).

24. Of course, this only raises the higher order question of what determines strategic understanding. Before playing our experiment, we asked the players for their academic major, the number of semesters they had completed at university, and their final secondary school grade (known in Germany as the Abitur). We also asked them questions to get an indication of their risk preferences, degree of trust, beliefs about others pursuing their self-interest, and reliance on intuition when making decisions. After that we let them play an incentivized "beauty contest game" in order to obtain a measure of their strategic sophistication
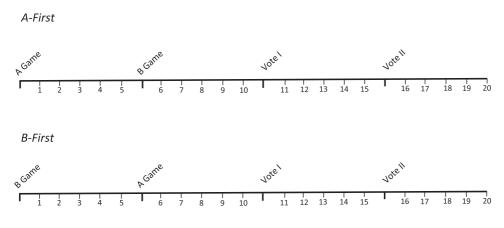
FIGURE 7. Play-First treatments.

4.5. *Play-First Treatment*

We have so far demonstrated that there exists both a push and a pull effect. We know that the push effect is necessary (and that the pull effect alone is not sufficient) because no group chose B without first trying A. Here we report the results of a new treatment. This shows that the pull effect is also necessary in order for players to choose B over A. Analysis of this new treatment also provides further evidence for the push effect.

In the new treatment, the players must choose between A and B-8. In contrast to *Vote-First-B-8*, however, the players must have experience playing both games before choosing. In one version of the new treatment, *A-First*, the players must play the A game in the first phase and the B game in the second phase. After that, they play the same way as in the *Vote-First* treatments, voting and then playing five contribution rounds in the third phase, and then repeating this sequence in the fourth and final phase. In the other version of the new treatment, *B-First,* the players must play B followed by A before voting in third and fourth phases; see Figure 7.

In *Vote-First*, the players must discover for themselves which game is best to play without the benefit of experience—a situation that comes closest to the real world examples we discuss in Section 5. The reason for the *Play-First* treatment is to see if and how behavior changes when players have experience playing both games before voting. By comparing this treatment with *Vote-First* we can thus determine how expectations in both games affect group behavior. By having the players play A first

---

(Bosch-Domènech et al. 2002). In contrast to Dal Bó, Foster, and Putterman (2010), however, we did not find any significant correlations between the personal characteristics and preferences of the players or the numbers they chose in the beauty contest game and the way these individuals voted in our *Vote-First* treatments. Finally, we tested whether a player's contribution decision in the very first round of the first phase of playing A was correlated with how this person voted. We found no statistically significant effects. A plausible interpretation of these results is that voting was determined by expectations, and that expectations could not be predicted from these elicited variables.

TABLE 5. *Play-First* treatments.

Phase	Game	Play-A-First-B-8			Play-B-First-B-8		
		% of groups	Average % contributions	Average payoff	% of groups	Average % contributions	Average payoff
I	A	100	32	6.6	0	–	–
	B	0	–	–	100	80	6.8
II	A	0	–	–	100	50	7.5
	B	100	73	6.3	0	–	–
III	A	60	24	6.2	60	23	6.2
	B	40	100	8.0	40	100	8.0
IV	A	40	9	5.5	30	16	5.8
	B	60	99	7.9	70	99	7.9

followed by B, or B first followed by A, we can also determine whether the order of experience has a separate effect from the experience itself.

As shown in Table 5, we do not find significant differences between *A-First* and *B-First* as regards how groups vote beginning in the third phase (Fisher's exact test, $p > 0.10$ for each phase) or the contributions they make following these votes, conditional on their choice of A or B (MWW test, $p > 0.10$ for each phase). We thus pool the data for both versions and call the combined treatment *Play-First-B-8*. The results for this combined treatment are shown in Figure 8.

Our focus is on whether the outcomes observed in the first two phases of *Play-First*, when all groups are required to play both A and B precisely once, affect the choice of which game to play in the second two phases. We are also interested in knowing how the choices made in these two voting phases compare with the choices made in the first two voting phases of *Vote-First*.

Before turning to these questions, we should note that contributions and payoffs, conditional on the game that has been chosen, reflect a similar pattern as before. As in *Vote-First-B-8*, the groups that chose to play B at the start of the third phase of *Play-First-B-8* contributed significantly more than the groups that chose to play A (MWW test, $p < 0.01$ for each phase). They also got a significantly higher payoff ($p < 0.01$ for each phase).

The important difference between *Vote-First-B-8* and *Play-First-B-8* lies in the choice of which game to play in the first two voting phases (phases one and two for *Vote-First-B-8* and phases three and four for *Play-First-B-8*). Only two out of 20 groups chose to play B at least once in the first two phases of *Vote-First-B-8*, whereas 15 out of 20 groups chose to play B at least once in the two voting phases of *Play-First-B-8* (Fisher's exact test, $p < 0.01$ for each phase). We infer from this evidence that the contrast in behavior between the two treatments reflects a difference in expectations (with this difference being shaped by behavior in the nonvoting phases of *Play-First*).

We can also compare behavior in the third and fourth phases of the *Play-First* and *Vote-First* treatments. In this comparison, groups in both treatments have gained the same amount of experience. The difference is that, in *Play-First*, this experience was

Blue = A Orange = B
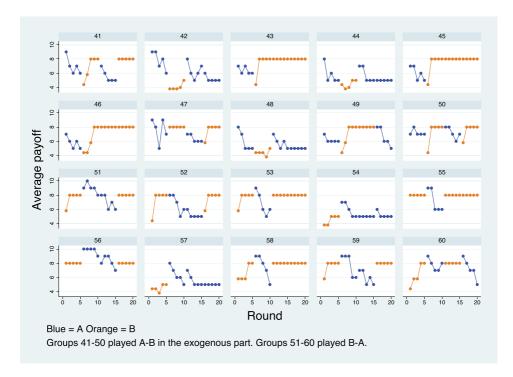Groups 41-50 played A-B in the exogenous part. Groups 51-60 played B-A.

FIGURE 8. Payoffs over time by group for *Play-First-B-8*.

imposed upon the groups, whereas in *Vote-First*, it was chosen by the groups. In the latter cases, this experience was also, for most groups, incomplete (that is, the groups had experience playing only one game). In this comparison, as with the previous one, more groups chose the B game in *Play-First* than in *Vote-First* (8 vs. 4 in the third phase and 13 vs. 9 in the fourth phase) but the differences are smaller and lack statistical significance (Fisher's exact test, $p > 0.05$ each). This shows that both kinds of experience help but that having experience playing both games seems to help a little more.

The surprise, perhaps, is that *any* group would choose A in the voting phases of *Play-First-B-8*. However, there were five instances of coordination failure in the nonvoting phases of *Play-First* (see Figure 8, groups 42, 44, 48, 54, and 57), an outcome never observed in *Vote-First*. The reason for this failure is probably due to the players being denied any opportunity to signal their intentions by voting.[25] As noted in Section 3, the prospects of players being able to coordinate on the mutually preferred equilibrium in B-8 are unclear in the absence of vote-signaling, as the inefficient equilibrium is risk-dominant in this game and the players could use risk-dominance as a selection rule. When coordination on the mutually preferred equilibrium failed in the

---

25. Behavior may also have been affected by the players not choosing for themselves which game to play. See Dal Bó, Foster, and Putterman (2010) and Sutter, Haigner, and Kocher (2010).

nonvoting phases, groups always chose to play A in the voting phases. Chastened by their bad experience playing B, these groups never attempted to play B again. Indeed, failure to coordinate on the mutually preferred equilibrium in the first two phases of *Play-First* is perfectly correlated with whether or not groups try game B *at all* in the last two phases (Spearman's $\rho = 1.00$, $p < 0.01$). A bad experience when playing B made these groups pessimistic about the prospects of coordination succeeding, squelching the pull effect.

Of course, the reasoned argument of Section 3 for the use of voting as a signaling device for coordination should not be affected by the way the game was played previously, in the phase in which groups were made to play B. Once given the opportunity to vote for which game to play, the players should have understood that the rules of the game had changed, and adjusted their expectations—and, hence, their behavior—accordingly. The fact that the groups that failed to coordinate when made to play B did not subsequently vote for B suggests that a majority of the members of these groups conditioned their expectations on the way their group members played in the past rather than thinking strategically about the future.

Importantly, we also find that the groups that failed to coordinate on the mutually preferred equilibrium in the first two phases of *Play-First* performed poorly when playing A (see Figure 8). Their average contribution rate over the last two phases of playing A is just 8%.[26] These players surely were under no illusions about cooperation in the A game, but they were pessimistic about the prospects of coordination succeeding in the B game. This demonstrates that the push effect is only a necessary and not a sufficient condition for switching, and that the pull effect is also necessary.

Apart from the five groups that failed to coordinate in the nonvoting phases of *Play-First*, only two other groups (49 and 60) played A in the final phase of this treatment. The behavior of these groups resembles that of group 28 in *Vote-First-B-8* (see Figure 5). These groups probably voted for A in the final voting phase believing or hoping that their contributions, which were high when they played B previously, would remain high if they switched to A, yielding them a larger payoff. We will never know, but it seems that these groups probably regretted this last switch, and that they would have chosen differently had they to do over again.[27] In any event, it's clear that the main difference between *Vote-First* and *Play-First* consists in the cases in which coordination failed. When coordination succeeded in the B game, making the players optimistic about the prospects for coordination, groups chose B over A. When coordination failed in the B game, making the players pessimistic about the prospects for coordination, groups chose A over B.

---

26.   Recalling the result from *Vote-First-B-8* that behavior in B is unrelated to previous behavior in A (Table 3), this result for *Play-First* suggests that behavior in the A game may be more sensitive to history than behavior in the B game.

27.   In the ex post questionnaire, nine out of the ten students in these two groups recommended that a new group play B; only one player recommended A. These responses lend support to our hypothesis that these groups would have switched back to B if given one more opportunity.

TABLE 6. A-voters in *Vote-First-B-8*.

Phase	Game	Number of groups	Average number of A-voters per group	Average A-voter % contributions	Average A-voter payoff
I	A	20	4.5	39	7.0
	B	0	–	–	–
II	A	18	3.8	23	6.3
	B	2	2	85	7.4
III	A	16	3.5	22	6.4
	B	4	1.5	85	7.1
IV	A	11	3.4	8	5.6
	B	9	1.3	85	7.3

Note finally that *Play-First* also provides more evidence of the push effect. There is a strong correlation between the average contribution level in the A game when played in the nonvoting phase and the voting phase in which a group chose to play B rather than A (Spearman's $\rho = 0.53$, $p = 0.04$). Groups that performed poorly when playing A in the nonvoting phase chose to play B at the first opportunity. Groups that performed better when playing A in the nonvoting phases needed to play A in another (frustrating) phase before switching to B.

4.6. The Difference Between A-Voters and B-Voters

Our main result—that about half of the groups in our *Vote-First-B-8* treatment never switched to B—stands in stark contrast to previous studies that found a gradual but eventually nearly complete movement to the superior regime (Ertan et al. 2009; Dal Bó et al. 2013). Did these groups make a mistake? Would they have done better by playing the B game? It is impossible to tell. On the one hand, every group that switched to B changed its behavior and did better. On the other hand, the groups that voted for A did so because they had different expectations for how they would behave when playing B, and expectations are of crucial importance to how these games are actually played.

Though we cannot be sure how the groups that played A would have done if they had played B, we can compare the payoffs of the A-voters who got their way, and were able to play A because of how their co-players voted, with the payoffs of the A-voters who found themselves in a minority, and were thus forced to play B. Two pieces of evidence show that the A-voters who played A earned less than the A-voters who were forced to play B.

The first piece of evidence comes from between-group comparisons. Table 6 compares the A-voters who played A with the A-voters who were forced to play B (because of how their fellow group members voted) for each phase. It shows that A-voters always earned more when playing the B game than when playing the A game. The differences in between-group behavior within the same phase are not statistically significant in the second and the third phases, but note that we have only few observations for these two phases. In the fourth phase, where we have

more observations, the difference is highly significant (MWW test, $p < 0.01$).[28] The differences become even more pronounced when we look at the A-voters' contributions under the two regimes. The A-voters who played B always contributed more than the A-voters who played A, and these differences are statistically significant ($p < 0.05$) in every phase.

The second piece of evidence comes from within-group comparisons. Ten groups switched from A to B at some point (we ignore here the second switch from A to B by groups 21 and 25; see Figure 5). Comparing the payoffs of the players who voted for A in both phases, before and after their group switched to B, we find that 69% of these (same) A-voters earned strictly more when they were forced to play B because of how their group voted. A Wilcoxon signed-rank test that takes the group average per phase as the unit of observation shows that the A-voters who got their way and were able to play A earned a lower payoff than when they were in a minority and were forced to play B ($p < 0.05$). Here, also, the differences become more pronounced when we look at contributions under the two regimes. In fact, *all* the A-voters contributed more when playing B than when playing A ($p < 0.01$).

This last result shows that the A-voters who had to play B were not only grouped with co-players who behaved differently, but that they themselves behaved differently under the two regimes. An interesting question is *when* these A-voters changed their behavior. It could be that, once B had been chosen, the A-voters thought more carefully about this game, noticed that voting to play this game signaled intentions, and saw that it was now in *their* best interest to contribute. Such a result would suggest that these individuals supported the A game before because they had not thought carefully enough about how others would play the B game or how they themselves would play this game. Indeed, Dal Bó et al. (2013) found that a significant share of subjects do not anticipate how their own behavior will change in the modified game. An alternative hypothesis is that the A-voters changed their behavior only in later rounds, after they had observed how the B-voters contributed. Such a behavior might best be described as "adaptive" rather than "strategic."[29]

The regression model in the second column of Table 3 shows that, after a group switches from playing A to playing B, the players who voted for A and lost were less likely to contribute in the first round of the B game than the voters who voted for B and won. In numbers: Out of the 17 A-voters who lost the vote, 12 made the same contribution decision in the two rounds just before and after the switch. Only four A-voters changed their decision from not contributing to contributing and one individual

---

28.   A more conservative comparison would include only the groups that have either two or three A-voters. These groups differ by only one A-vote, and so may be less likely to differ in some unobservable ways. The results remain essentially the same. In all phases, A-voters earned more when playing B than when playing A. The difference in earnings in the fourth phase remains significant (MWW test, $p = 0.01$). In each phase, A-voters in B contributed more than A-voters in A, and all the differences are statistically significant ($p < 0.05$ each).

29.   As noted in footnote 11, strategic and adaptive behaviors are explored in different strands of the theoretical literature on equilibrium selection.

changed from contributing to not contributing. The B-voters behaved very differently. Of the 33 B-voters who got their way in the vote, 27 changed their contribution decision after the switch from not contributing to contributing, five did not change their decision, and one changed from contributing to not contributing. These results suggest that the majority of A-voters were initially pessimistic about the prospects of coordination succeeding in the B game but became more optimistic over time. By the end of each phase of playing B, they all contributed. By contrast, the majority of B-voters seemed to be optimistic about the prospects of coordination succeeding from the very beginning.

As discussed in Section 3, ambiguity about the prospects of coordination succeeding in B-8 should be resolved by vote signaling. It thus appears that the A-voters who did not contribute when B was chosen might have failed to read this signal. This failure can also help to explain why these people voted for A in the first place. However, when the A-voters were out-voted and had to play B, they learned that they do better by contributing, since *all* groups coordinated flawlessly by the end of every phase in which B was played. Eight of the ten groups that played B reached this point of flawless coordination in the third round at the latest (see Figure 5). As a consequence, and as mentioned before, when we compare behavior at the phase-level, A-voters contributed and earned more when they were forced to play B than when they played A. Unfortunately, the A-voters who were in the majority never had the chance to learn how to behave when playing B.

5. Applications

In this section we show how our experimental results can be helpful for interpreting three real world examples of international agreements adopting different approaches.

We begin with the International Convention for the Prevention of Pollution by Ships, more commonly known as MARPOL. MARPOL establishes a technology standard for oil tankers, ensuring that a tanker's oil cargo is kept physically separate from its ballast water. Previously, most oil pollution in the oceans resulted from tankers flushing out their ballast water mixed with oil. Under MARPOL, however, port states can protect their coasts simply by restricting entry to tankers meeting the new standard—that is, by banning trade involving the old technology. As the global market for ocean shipping is characterized by strong network externalities, this technology-standards approach creates incentives for port states and tanker owners alike to adopt the new standard once assured that a critical mass of others will adopt the new standard. MARPOL thus made protection of the oceans a tipping game.[30]

However, choice of this approach came at a cost. The direct approach of limiting emissions was "cheaper, more economically efficient, and 'in theory… a good idea'"

---

30. For a theoretical model showing this kind of transformation, see Barrett (2006); see also Barrett (2003).

(Mitchell 1994, p. 434), but was difficult to monitor. The mandated technology-standards approach, by contrast, "was expensive both in terms of capital and the reduction to cargo-carrying capacity" (Mitchell 1994, p. 434), but was easy to monitor and so could be enforced. Today, virtually all oil tankers comply with the MARPOL standard. However, as in our *Vote-First-B-8* treatment, negotiators adopted MARPOL's coordination approach very reluctantly. They first sought to reduce discharges directly and they persisted in trying to make this approach work for more than 50 years. It was not until the 1970s that they switched to the technology-standards approach.

The Montreal Protocol on protecting the ozone layer works a little differently than MARPOL, but has had a similarly transformative effect. Montreal restricts both the consumption and production of chlorofluorocarbons (CFCs), while also banning trade in CFCs and products containing CFCs between parties and nonparties. Under Montreal, provided enough countries limit their consumption of CFCs, exporters want to produce the CFC substitutes; and provided enough countries produce the substitutes, importers want to limit their consumption of CFCs. Like MARPOL, Montreal's approach makes protection of the ozone layer a tipping game.[31] The important difference is that Montreal sustains an outcome that is indistinguishable from a first best. Rather than mandate a particular substitute (a technology standard), Montreal only mandates reductions in CFCs (a performance standard), leaving it to the parties ("the market") to choose which substitutes to employ. As in our *Vote-First-B-10* treatment, negotiators of the Montreal Protocol adopted the coordination approach right from the start.

Unlike our first two examples, the Kyoto Protocol on climate change typifies the direct approach to the prisoners' dilemma. Kyoto specifies national greenhouse gas emission limits without the support of an agreed enforcement mechanism.[32] When this approach was first put to the test, it crumbled. The United States refused to ratify the agreement, Canada withdrew from it, and Japan, New Zealand, and Russia decided not to participate in the Protocol's second phase. Although other countries—notably members of the European Union—have taken steps to reduce their emissions, overall the agreement has had little if any effect (Aichele and Felbermayr 2011). Interestingly, Kyoto incorporates several flexible implementation mechanisms including a provision allowing emissions trading. The people who negotiated Kyoto thus focused their attention on cost-effectiveness, not enforcement.

Like Kyoto, the Paris Agreement, adopted in late 2015, also reflects the direct approach to the prisoners' dilemma. The difference is that the national contributions pledged in Paris were chosen independently rather than negotiated and are explicitly recognized as being voluntary. It was little noticed at the time, but the same group of countries that met in Paris in December agreed a month before in Dubai to commence negotiation of an amendment to the Montreal Protocol to phase down

---

31. For a theoretical model of this transformation, see Barrett (1997, 2003). See also Heal and Kunreuther (2012).

32. Article 18 says that any compliance mechanism applying with "binding consequences" must be agreed by amendment, and no such amendment has been adopted.

hydrofluorocarbons, a chemical that does not destroy the ozone layer but that is a powerful greenhouse gas. Our research suggests that this amendment is to be welcomed, and that negotiators would do well to explore further opportunities for tipping, including second best approaches like technology standards combined with trade restrictions.

6. Conclusions

In many settings players can decide on the rules of the game before they begin playing the game. For example, when negotiators meet to adopt an international agreement to provide a public good, they must decide which game to play. A prisoners' dilemma can potentially achieve the overall first best outcome, but collective action in this game is difficult to enforce. Collective action is easier to enforce in a tipping game, but coordination in this game may not be assured and choice of this game may foreclose the possibility of attaining the first best.

   The problem with choosing between these games is that players cannot be certain which game will work best. Our experiment shows that players are quick to choose the tipping game when doing so enables them to sustain the overall first best outcome and when this outcome is risk-dominant. However, they are reluctant to choose this game when the efficient equilibrium in the B game is not risk-dominant and when doing so means settling for the second best, even if the second best outcome is better than the one that results when the players try, but fail, to cooperate in the prisoners' dilemma. In this case, many groups choose to play the prisoners' dilemma throughout the experiment, despite this game's strategic disadvantage and the difficulty these groups have in sustaining cooperation. We do not know if these groups would have been able to coordinate in the B game (were they to have played B). After all, if group members are pessimistic about coordination succeeding, failure in the B game will be a self-fulfilling prophecy. We do know, however, that it was beneficial for the A-voters to lose the vote. The A-voters who were out-voted and thus made to play B did better than the A-voters who were in the majority and got to play their preferred game. The loss gave these players an opportunity to update their expectations and learn about the strategic advantage of the tipping game. This was not possible for the A-voters who won the vote.

   Our results confirm the tendency observed in previous studies for players to misapprehend the consequences of the choice of which game to play (Dal Bó et al. 2013). However, this tendency is unusually striking and persistent in our experiment. In the treatment in which the tipping game can sustain only a second best outcome and there is a conflict between payoff- and risk dominance, *every* group started out by choosing the prisoners' dilemma. This game appears to be the default choice when players are unsure how the two games will be played.

   The previous literature has also found that, when given the opportunity to revise their choice of institution, players will gradually move toward the welfare improving institution (Gürerk, Irlenbusch, and Rockenbach 2006; Ertan, Page, and Putterman

2009; Dal Bó et al. 2013; Markussen, Putterman, and Tyran 2014).[33] Our results are different and more unsettling. We find that a significant number of groups remain loyal to the prisoners' dilemma even after they have witnessed their repeated failure to sustain much cooperation in this game. Over the course of our experiment, cooperation in the prisoners' dilemma deteriorated significantly, and yet only half of the groups switched to the tipping game.

Why do only some groups switch? To be willing to switch, groups not only had to become disillusioned with cooperation in the prisoners' dilemma; they also had to be hopeful about the prospects for coordination in the tipping game. The skill needed to anticipate other players' behavior in the two games is thus crucial. In particular, comparison of the *Vote-First* and *Play-First* treatments shows that awareness of vote-signaling behavior is critical to success in the B game—and, therefore, to the players' willingness to vote for B. Our interpretation of this result is that many voters seem oblivious of the signaling effect of voting. It remains for future research to show whether our results are unique to the game choice studied in our experiment or whether these results reflect a more general tendency for a significant fraction of players to misapprehend the meaning of signals.

Another question we leave for future research is whether the group members that choose to remain in the prisoners' dilemma could be made to see that they would do better by switching and whether, having come to see this, they would then act in ways that *ensured* they did better by switching. In other words, are there ways to change the possibly self-fulfilling expectations of an entire group of players?

Appendix: Experimental Instructions

Here we provide the instructions for the *Vote-First-B-10* treatment, translated from German. Instructions for the other treatments are available upon request.

Welcome to our experiment!

A.1. General information

In our experiment you can earn money. How much you earn will depend on the game-play, or more precisely on the decisions you and your fellow co-players make. For a successful run of this experiment, it is essential that you do not talk to other participants. Now read the following rules of the game carefully. If you have any questions, give us a hand signal. We will come to you and answer them.
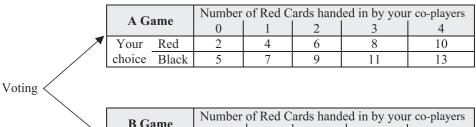
---

33.   For example, in a recent experiment on endogenous punishment institutions, Markussen, Putterman, and Tyran (2014, p. 303) found that "voters manage surprisingly well to self-organize for collective action, and (. . .) provide a remarkable example of efficient endogenous emergence of institutions."

A.2. Game rules

There are five players in your group, meaning you and four other players. Each player is faced with the same decision problem. All decisions are anonymous. For this reason, you will be identified by a number (between 1 and 5), which you will see in the lower left corner of your display.

There are two games, Game A and Game B. At the beginning, every player in your group will vote for one of the two games. After that, and before the game starts, the players' votes will be displayed to everyone. The game that receives the most votes (at least three out of five) will be played by the group. Thus, the group plays either Game A or Game B.

In each game, you will receive two cards, a Red Card and a Black Card. You will be asked to hand in one of the two cards. Your payoff will depend on which game is played (A or B), which card you hand in (Red or Black), and which cards your four co-players hand in. The following two tables show your payoff for all possible outcomes in each game.

A Game		Number of Red Cards handed in by your co-players				
		0	1	2	3	4
Your	Red	2	4	6	8	10
choice	Black	5	7	9	11	13

Voting

B Game		Number of Red Cards handed in by your co-players				
		0	1	2	3	4
Your	Red	2	4	6	8	10
choice	Black	5	5	5	5	5

Here are some examples for how to read the tables:

If the group plays the A game and two of your co-players hand in their red card (and the other two co-players hand in their black card), you get six tokens if you hand in your red card and you get nine tokens if you hand in your black card.

If the group plays the B game and two of your co-players hand in their red card (and the other two co-players hand in their black card), you get six tokens if you hand in your red card and you get five tokens if you hand in your black card.

The game (A or B) that receives the most votes in the group (at least three out of five) will be played five times consecutively. In each round you will be asked to hand in either the red card or the back card. After this, your group will vote again, play the chosen game another five times, and so on. In total, your group will vote four times and play the chosen game five times after each vote. Hence, you and your co-players will decide which card to hand in 20 times in total. You will play with the same group of players throughout all rounds. The sum of tokens you earn across all 20 rounds will be paid to you in cash at the end. You will get €0.10 for each token. For example, if you earn 150 tokens in total, you will get €15.00.

A.3. *Control questions*

Please answer the following control questions.

a. Right or wrong? At the beginning all players will vote for Game A or Game B. After everyone votes, and before the game starts, you will learn how your co-players voted and they will learn how you voted. The game that receives the most votes will be played by the group.

◯ Right   ◯ Wrong

b. Right or wrong? The group will vote four times in total. After each vote, the chosen game will be played for five rounds.

◯ Right   ◯ Wrong

c. Assume that the group plays the A Game and one of your co-players hands in the Red Card (and the other three co-players hand in their Black Card). What is your payoff if you hand in your Red Card? _____ What is your payoff if you hand in your Black Card? _____

d. Assume that the group plays the B Game and one of your co-players hands in the Red Card (and the other three co-players hand in their Black Card). What is your payoff if you hand in your Red Card? _____ What is your payoff if you hand in your Black Card? _____

e. Assume that the group plays the A Game and three of your co-players hand in their Red Card (and the other co-player hands in the Black Card). What is your payoff if you hand in your Red Card? _____ What is your payoff if you hand in your Black Card? _____

f. Assume that the group plays the B Game and three of your co-players hand in their Red Card (and the other co-player hands in the Black Card). What is your payoff if you hand in your Red Card? _____ What is your payoff if you hand in your Black Card? _____

g. Assume that the group plays the A Game and all four of your co-players hand in their Red Card (and no one hands in the Black Card). What is your payoff if you hand in your Red Card? _____ What is your payoff if you hand in your Black Card? _____

h. Assume that the group plays the B Game and all four of your co-players hand in their Red Card (and no one hands in the Black Card). What is your payoff if you hand in your Red Card? _____ What is your payoff if you hand in your Black Card? _____

Please also consider other examples! Give us a hand signal after you have answered all the control questions. We will come to you and check that you have answered all the questions correctly. The game will begin after we have checked the answers of all the participants and answered any questions you may have. Good luck!

References

Aichele, Rahel and Gabriel Felbermayr (2011). "Kyoto and the Carbon Footprint of Nations." *Journal of Environmental Economics and Management*, 63, 336–354.

Barrett, Scott (1997). "The Strategy of Trade Sanctions in International Environmental Agreements." *Resource and Energy Economics*, 19, 345–361.

Barrett, Scott (2003). *Environment and Statecraft: The Strategy of Environmental Treaty-Making*. Oxford University Press, Oxford.

Barrett, Scott (2006). "Climate Treaties and 'Breakthrough' Technologies." *American Economic Review (Papers and Proceedings)*, 96, 22–25.

Bosch-Domènech, Antoni, José G. Montalvo, Rosemarie Nagel, and Albert Satorra (2002). "One, Two, (Three), Infinity: Newspaper and Lab Beauty-Contest Experiments." *American Economic Review*, 92(5), 1687–1701.

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

Crawford, Vincent P. (1997). "Theory and Experiment in the Analysis of Strategic Interaction." In *Advances in Economics and Econometrics: Theory and Applications,* Vol. 1, edited by Kreps D.M. and K.F. Wallis. Cambridge University Press, Cambridge, pp. 206–242.

Croson, Rachel T. A. and Melanie Beth Marks (2000). "Step Returns in Threshold Public Goods: A Meta- and Experimental Analysis." *Experimental Economics*, 2, 239–259.

Dal Bó, Pedro (2014). "Experimental Evidence on the Workings of Democratic Institutions." In *Economic Institutions, Rights, Growth, and Sustainability: The Legacy of Douglass North*, edited by Galiani Sebastian and Itai Sened. Cambridge University Press, Cambridge.

Dal Bó, Ernesto, Pedro Dal Bó, and Erik Eyster (2013). "The Demand for Bad Policy When Voters Underappreciate Equilibrium Effects." Working paper, Haas School of Business, University of California, Berkeley.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman (2010). "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review*, 100(5), 2205–2229.

Devetag, Giovanna and Andreas Ortmann (2007). "When and Why? A Critical Survey on Coordination Failures in the Laboratory." *Experimental Economics*, 10, 331–344.

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

Fischbacher, Urs (2007). "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10, 171–178.

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

Heal, Geoffrey and Howard Kunreuther (2012). "Tipping Climate Negotiations." In *Common Sense and Climate Change: Essays in Honor of Thomas Schelling*, edited by Hahn Robert and Alistair Ulph. Oxford University Press, Oxford.

Hufbauer, Gary Clyde, Jeffrey J. Schott, Kimberly Ann Elliott, and Barbara Oegg (2007). *Economic Sanctions Reconsidered*, 3rd ed. Institute for International Economics, Washington, DC.

Ledyard, John O. (1995). "Public Goods: A survey of Experimental Research." In *Handbook of Experimental Economics*, edited by Kagel John and Alvin Roth. Princeton University Press, Princeton.

Markussen, Thomas, Louis Putterman, and Jean-Robert Tyran (2014). "Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes." *Review of Economic Studies*, 81, 301–324.

Mitchell, Ronald B. (1994). "Regime Design Matters: Intentional Oil Pollution and Treaty Compliance." *International Organization*, 48, 425–458.

Schelling, Thomas C. (1978). *Micromotives and Macrobehavior*. W.W. Norton, New York.

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

Supplementary Data

Supplementary data are available at *JEEA* online.