Columbia | SIPA
Center for Environmental Economics and Policy

CEEP Working Paper Series
Working Paper Number 3

April 2019

Coercive Trade Agreements for Supplying Global Public Goods

Scott Barrett, Astrid Dannenberg

Coercive Trade Agreements for Supplying Global Public Goods

Scott Barrett[*]
Columbia University

Astrid Dannenberg[+]
University of Kassel

7 April 2019

Abstract. International cooperation on trade is less vulnerable to free riding than international cooperation to supply a global public good like climate change mitigation. Can linking cooperation on these issues increase welfare? Simple theory shows that, at best, linkage is a coordination game in which equilibrium selection is unreliable. In the experimental lab we find that whether linkage helps or hurts depends on the gains from cooperating on trade relative to the gains from cooperating to supply the public good and the institutional setting. Groups do better when the relative gains are high and decisions are made multilaterally rather than unilaterally.

[*]sb3116@columbia.edu. [+]dannenberg@uni-kassel.de

Acknowledgements. We are grateful for the comments we received from participants attending the workshop on Linkage Diplomacy: Climate Change and International Trade, held at Columbia University on 23 February 2018. This research was financially supported by the European Union (EU) Horizon 2020 program, action ERC-2014-STG, Project HUCO, grant number 636746.

“Let’s sign no more trade agreements with powers that don’t respect the Paris Agreement [on climate change].”

Emmanuel Macron, President of France[1]

1. Introduction

International cooperation to liberalize trade is fundamentally easier than international cooperation to supply a global public good. As trade is bilateral, a multilateral trade agreement specifies behavior among a *set* of *two-player* prisoners’ dilemma games, allowing cooperation to be sustained by *direct reciprocity*.[2] Global public goods, by contrast, are supplied by agreements specifying behavior in a *single N-player* prisoners’ dilemma game, requiring cooperation to be sustained by *diffused reciprocity*. Given this difference, countries may be tempted to link cooperation on trade to cooperation in supplying a global public good. In this paper we inquire into the opportunities and risks associated with this decision.

There are many public goods that might be linked to trade, including human rights, nuclear non-proliferation, respect for intellectual property rights, and environmental protection (Maggi 2016), though the candidate for linkage that has attracted the most interest so far is climate change mitigation.[3] In his Presidential Address to the American Economic Association, William Nordhaus (2015) offered an analysis of the consequences of a coalition of likeminded countries adopting a common carbon tax, and then imposing a tariff on imports from non-members.[4] In his paper, all countries

[1] Speech to the United Nations General Assembly, 25 September 2018.
[2] A multilateral trade agreement comprising N members thus specifies behavior for country pairs.
[3] None of the climate agreements negotiated thus far establish a link to trade, but an early draft of the Paris Agreement considered a range of potential options, from explicitly avoiding mention of “unilateral measures in the agreement” to incorporating such measures “upon Parties’ request.” The WTO’s position is currently ambiguous, though rulings on adjudicated disputes suggest that the WTO would view multilateral efforts more favorably than unilateral ones.
[4] Lessmann, Marschinski, and Edenhofer (2009) analyzed this same situation previously and came to similar conclusions. Böhringer, Carbone, and Rutherford (2016), using a CGE model, analyze the effect

have an incentive to free ride by not adopting the carbon tax, but each country also wants to avoid being subject to a punitive tariff. Under some circumstances, Nordhaus shows, a coercive coalition can support universal membership, meaning that the global public good is supplied without trade actually being restricted. Under other circumstances, he finds, linkage sustains only partial cooperation (meaning that only coalition members supply the global public good and that trade between these countries and non-members is restricted) or no cooperation on climate change (in which case trade is unaffected).

Nordhaus calls the institution that links climate change to trade a "climate club." Here, we call it a *Coercive Trade Agreement*, a label that invites ready comparison to the much-studied preferential trade agreement.[5] Members of a preferential trade agreement lower the tariffs they impose on one another without increasing the tariffs they impose on non-members, whereas members of a coercive trade agreement raise the tariffs they impose on non-members without changing the tariffs they impose on one another.[6] Preferential trade agreements may harm non-members (due to trade diversion), but such an effect is incidental to the purpose of these agreements. Coercive trade agreements, by contrast, are intended to harm non-members, the aim being to coerce these countries into joining the coalition to supply the public good.

A critical assumption underpinning Nordhaus's analysis is that, before the linked game is played, all countries consent to be bound by an agreement to prohibit non-members from retaliating against the climate coalition (Nordhaus 2015). This assumption is critical because a country will only refrain from imposing tariffs against

of a coalition of countries imposing a carbon tariff against non-coalition members. For related analyses, see Böhringer and Rutherford (2017) and Winchester (2018).

[5] For an historical perspective on preferential trade agreements, see Krueger (1999); for a collection of analytical papers, see Bhagwati, Krishna, and Panagariya (1999).

[6] A less controversial form of linkage would involve members of a coalition lowering the tariffs they impose on one another, while leaving the tariffs they impose on non-members unchanged. However, as today's tariff levels are already low, the gain in real income from lowering tariffs is likely to be a lot smaller than the absolute value of the loss in real income that would be caused by increasing tariffs (Ossa 2014). Today, raising tariffs offers more leverage for supplying a global public good.

others if it believes that this behavior will be reciprocated (Bagwell and Staiger 1999). A promise not to retaliate thus lacks credibility.

In this paper we present a simple model that exposes both the opportunities and the risks associated with linkage. We show that the nature of the linked game depends on key parameter values—basically, how the gains from mutual reductions in tariffs compare to the gains from free riding in the supply of the public good. When retaliation is prohibited, the linked game could be a cooperation game in which full cooperation is the unique Nash equilibrium; a coordination game with two pure strategy Nash equilibria, only one of which is Pareto efficient; a chicken game in which some countries cooperate and some do not cooperate in equilibrium; or a prisoners' dilemma. By contrast, when retaliation is permitted, the linked game can only be a prisoners' dilemma or a coordination game. If the linked game is a prisoners' dilemma, linkage has no advantage. By contrast, if the linked game is a coordination game, linkage is collectively desirable *but only if coordination on the efficient equilibrium succeeds*. We show that the linked game is more likely to be a prisoners' dilemma when retaliation is allowed than when it is prohibited. Moreover, even when the linked game is a coordination game when retaliation is permitted, we show that the payoff and risk dominant Nash equilibria differ, implying that coordination on the efficient equilibrium cannot be taken for granted.[7] At best, linkage is risky.

This analysis assumes that countries choose whether or not to link unilaterally. Unilateralism is the default rule in international relations, but it is not the only option. As recognized by Nordhaus (2015: 1351-2), though a linked agreement might emerge from spontaneous, bottom-up behavior, it might also be negotiated at "a grand Bretton-Woods-type conference." To explore linkage under a regime of multilateralism, we construct a stage game in which the players that participate in an agreement choose to supply the public good and impose tariffs on nonparticipants,

[7] For recent discussions of the theoretical challenges posed by selection, see Crawford (2016) and Samuelson (2016).

but only if the level of participation exceeds the tipping point for the linked game, in which case, every non-party to the agreement does better by joining. By design, the treaty thus *assures* selection of the Pareto efficient outcome. Note that supply of global public goods like protection of the ozone layer and prevention of oil pollution at sea were achieved by international agreements reflecting a similar design.[8]

Under a regime of unilateralism, coordination is unreliable, making linkage risky. Under a regime of multilateralism, coordination is assured, but only if the players see the strategic advantage in coordination and choose to link the different issues. Nordhaus (2015) solves for an equilibrium using an evolutionary computer algorithm, concealing these tensions. Here we explore behavior under both regimes in the lab. In our experiment, the choice of whether or not to link is determined by a vote. In *Unilateral*, so long as at least one player votes to link, all players must play the linked game. In *Multilateral*, an agreement to link only enters into force if a majority of players votes to link, with the size of the majority determined by the tipping point for the linked game.

We also explore sensitivity of this behavior to whether the gains from cooperation on trade are high or low relative to the gains from cooperation in supplying the public good (*High* versus *Low*). In theory, this difference should have no effect on behavior. However, in the *High* treatments, the tipping point for coordination is relatively low, and in the Low treatments the tipping point is relatively high, and behavior may depend on these differences.

[8] The Montreal Protocol to protect the ozone layer bans trade in CFCs between parties and non-parties, creating a tipping situation, as illustrated in the model by Barrett (1997), with coordination being achieved by a minimum participation level determined with reference to the tipping point. Consistent with this interpretation, Wagner (2015) demonstrates that, because of the trade restriction, early adoption of the agreement by some countries caused others to adopt it, making joining a strategic complement. The international agreement to limit oil pollution at sea, known as MARPOL, similarly achieved coordination in the adoption of a new technical standard for oil tankers, thanks to strong network externalities and the incentive coastal states had to block non-complying tankers from gaining access to their ports (Barrett 2003).

Our experiment reveals a tradeoff. When players choose unilaterally, linkage is more likely to be attempted, but coordination is less likely to succeed. When players choose multilaterally, linkage is less likely to be attempted but coordination is more likely to succeed. Despite this tradeoff, we find that, in welfare terms, multilateralism outperforms unilateralism, especially when the gains from trade cooperation are low relative to the gains from cooperating to supply the global public good.

The remainder of our paper is organized as follows. In the next two sections we present our model of cooperation on trade, followed by our model of cooperation in supplying a global public good. In Section 4 we characterize the game in which cooperation on trade is made conditional on cooperation in supplying the public good, assuming that retaliation is prohibited; and in Section 5 we perform the same exercise, assuming that retaliation is permitted. These sections assume that the decision to link is made unilaterally. In Section 6 we present a model in which linkage is chosen multilaterally. In Section 7 we describe our experimental design, and in Section 8 we present and discuss the main results of our experiments (less important results are shown in an Online Appendix). The final section summarizes our main findings and highlights priorities for further research.

2. Cooperation on trade

We begin by presenting a model of trade cooperation in which an agreement to set zero tariffs is self-enforcing.[9] Suppose there are N identical countries, that each country trades with every other country, and that every country *i* can impose a tariff $t_i^j \in \{0,t\}$, $t > 0$, against any country j, $j \neq i$. With $\pi_i^j(t;0) > \pi_i^j(0;0) > \pi_i^j(t;t) > \pi_i^j(0;t)$, each player has a dominant strategy to impose the tariff, but every pair of countries *i* and *j* is better off collectively when both countries show restraint.

[9] The assumption of zero tariffs is made only for simplicity. Our model is consistent with the World Trade Organization's (2009: 21-22) description of the "traditional approach" to trade agreements.

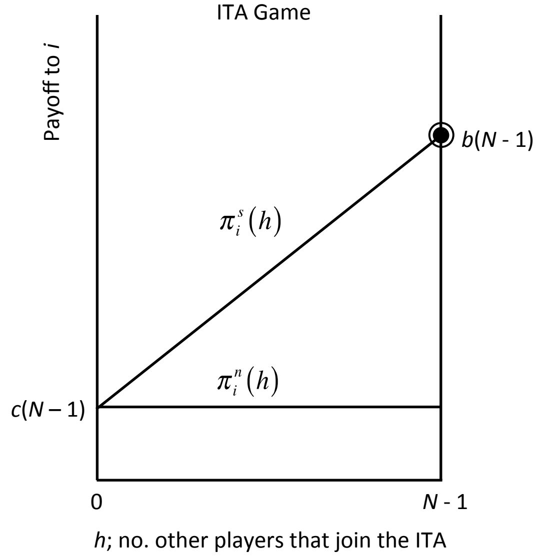
Wanting to overcome the prisoners' dilemma in tariff setting, suppose that all players play an international trade agreement (ITA) game in three stages. In stage 1, each player exercises its sovereignty by choosing independently whether or not to join the trade agreement. In stage 2, taking these participation decisions as given, members of the ITA choose their reciprocal tariffs collectively. Finally, in stage 3, taking as given all of the decisions made previously, ITA members choose independently whether to impose tariffs against non-members and non-members choose independently whether to impose tariffs against every other country, including members of the ITA. Solving the game backwards, it is obvious that: (i) in stage 3, each non-member will impose a tariff $t > 0$ against all other countries and each ITA-member will impose the same tariff $t > 0$ against all non-members; (ii) in stage 2, all members will impose a tariff of zero against one another; and (iii) finally, in stage 1, all countries will choose to join the ITA.

Let us now be more specific and assume that the payoffs are as follows: $\pi_i^j(t;0) = a$, $\pi_i^j(0;0) = b$, $\pi_i^j(t;t) = c$, $\pi_i^j(0;t) = d$, with $a > b > c > d$.[10] Solving the above game backwards, and focusing on the stage 1 game, suppose that h other countries have joined (or are expected to join) the ITA. Then, country i will choose to be a signatory (denoted by the superscript s) rather than a non-signatory (n) so long as $\pi_i^s(h) = hb + (N-h-1)c \ \geq \ \pi_i^n(h) = (N-1)c$; otherwise i will choose to be a non-signatory. The game is illustrated in Figure 1, in the style developed by Schelling (1978). Obviously, full participation in this game is a (strict) Nash equilibrium (indicated by the closed dot in the figure). Our model thus predicts that all players will join the ITA and that the regime will sustain universal free trade (the full cooperative outcome is indicated by the open circle).

[10] Parameters a-d and tariff levels 0 and t are specified exogenously here, but they can be thought of as representing solutions to optimization exercises in which payoff functions are maximized by choice of tariffs along a continuum. From this perspective, for example, the tariff t represents the symmetric Nash equilibrium tariff, and d the payoff associated with this tariff. For an example of this kind of calculation, see Ossa (2011). Note as well that the tariff level 0 can be interpreted as reflecting a normalization of the full cooperative tariff level.

Fig. 1

ITA Game



3. Voluntary cooperation in supplying the global public good

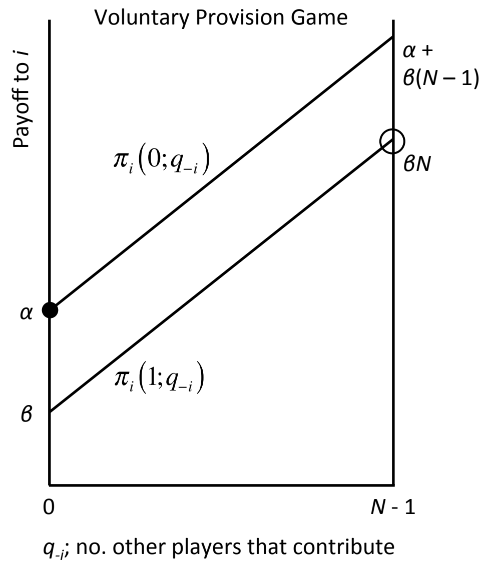
In keeping with our approach to modeling trade, suppose that every country *i* must choose whether or not to contribute to the global public good. Specifically, assume: $q_i \in \{0, q\}$, $q > 0$, with $\pi_i(0; q_{-i}) > \pi_i(q; q_{-i})$, where $q_{-i} = \sum_{j=1, j\neq i}^{N} q_j$, and $\pi_i(q; (N-1)q) > \pi_i(0;0)$. This is the classic public goods game.

We can be more specific and assume that payoffs are linear, with $\pi_i(0; q_{-i}) = \alpha + \beta q_{-i}$ and $\pi_i(q; q_{-i}) = \beta(q + q_{-i})$. The above conditions will then be satisfied provided $\beta N > \alpha > \beta > 0$. In keeping with Nordhaus's model, and with the design of the Paris Agreement on climate change, we assume that contributions to the public good are purely voluntary.[11] The climate change game is shown in Figure 2. For every player,

[11] For an experimental analysis of Paris, see Barrett and Dannenberg (2016). Note that we could model cooperation in supplying the global public good as being achieved by an international environmental agreement, making cooperation a chicken game; see Barrett (1994, 2003), Carraro and Siniscalco (1993), and Finus (2001); and for analyses incorporating trade, see Barrett (1997) and Eichner and

irrespective of how others choose, not contributing to the global public good yields a higher payoff than contributing, and yet all players are better off collectively when every player contributes.

Fig. 2



4. Linkage with retaliation prohibited

Here we model the effects of making cooperation on trade conditional on cooperation in supplying the public good, assuming (like Nordhaus 2015) that non-members of the coercive trade agreement (CTA) are prohibited from retaliating. In this model, signatories to the agreement supply the public good, trade freely with the other members, and impose a tariff t > 0 upon all non-signatories. Non-signatories, by contrast, free ride on signatories' provision of the public good, trade freely with other non-signatories, and (by assumption) refrain from imposing a positive tariff on signatories.

Pethig (2013). However, such self-enforcing agreements typically improve very little on the Nash equilibrium of the voluntary provision game. Using the parameter values in our experiment $(\alpha = 5, \beta = 2)$, for example, it is easy to show that the self-enforcing agreement would comprise three players for any $N \geq 3$. For all of these reasons, and to keep our experiment simple, the assumption of voluntary provision seems reasonable.

With these assumptions, payoffs in the linked game become $\pi_i^s = \beta(k+1) + bk + a(N-k-1)$ and $\pi_i^n = \alpha + \beta k + b(N-k-1) + dk$, where the superscripts *s* and *n* denote that the player is a signatory or a non-signatory to the coercive trade agreement, and *k* represents, from player *i*'s perspective, the number of other countries that participate in the agreement.

Setting $\pi_i^s = \pi_i^n$ gives

$$\tilde{k} = \frac{\left[(\alpha-\beta)-(a-b)(N-1)\right]}{\left[(b-d)-(a-b)\right]}. \tag{1}$$

The above payoffs imply:

Proposition 1. *Depending on parameter values (see Table 1), and assuming that retaliation against members of the coercive trade agreement is prohibited, the linked game is either: (i) a prisoners' dilemma in which membership in the CTA is empty; (ii) a chicken game in which membership in the CTA is partial (and equal to the smallest integer greater than* $\tilde{k}$ *); (iii) a coordination game with tipping point* $\tilde{k}$ *in which there are two (pure strategy) Nash equilibria, one that sustains zero participation and another that sustains full participation; or (iv) a cooperation game in which membership in the CTA is universal. When the CTA is a coordination game the payoff-dominant Nash equilibrium is risk dominant if and only if* $(\alpha-\beta)/(a-d) > (N-1)/2$.

Table 1. Properties of Linked Game: Retaliation Prohibited

Condition 1	Condition 2	Game
$(\alpha-\beta) > (a-b)(N-1)$	$(\alpha-\beta) > (b-d)(N-1)$	Prisoners' dilemma
$(a-b)(N-1) > (\alpha-\beta)$	$(\alpha-\beta) > (b-d)(N-1)$	Chicken
$(\alpha-\beta) > (a-b)(N-1)$	$(b-d)(N-1) > (\alpha-\beta)$	Coordination
$(a-b)(N-1) > (\alpha-\beta)$	$(b-d)(N-1) > (\alpha-\beta)$	Cooperation

To interpret Table 1, note that $(\alpha-\beta)$ represents a player's gain from not supplying the global public good,[12] $(a-b)(N-1)$ represents a player's gain from imposing tariffs upon all others when this player is the only member of the CTA, and $(b-d)(N-1)$ represents a player's gain to avoiding the imposition of tariffs by all others when this player joins the others in the CTA.

In a prisoners' dilemma, the gain to free riding exceeds both of the other gains, making not joining the CTA a dominant strategy. In a cooperation game, the reverse is true: trade cooperation dominates cooperation in supplying the public good, making joining the CTA a dominant strategy. In a chicken game, a player does better by being a member of the CTA when no other state is a member but it does worse by being a member when all other states are members. Finally, in a coordination game, a player does worse by being a member of the CTA when no other state is a member and does better by being a member when all other players are members. In this last situation, there are two Nash equilibria (in pure strategies), only one of which is efficient.[13]

Nordhaus's (2015) numerical results show that participation in a climate club may be full, partial, or empty, depending on the carbon tax and tariff combination (see in particular his Figure 6). Our model not only reveals these same equilibria, but also describes the games that give rise to them. In particular, we can't tell from Nordhaus's analysis whether, when all countries find it advantageous to join the club, the reason is that the underlying game is a cooperation game or a coordination game. Nor can we tell from his analysis whether, when the linked game is a coordination game, the

---

[12] In linear public goods experiments, β/α represents the marginal per capita return (MPCR) to supplying the public good. The literature (see Ledyard 1995) shows that contributions to a public good increase in the MPCR, suggesting that the need for enforcement increases as $(\alpha-\beta)$ increases.

[13] To sharpen the presentation, Table 1 (by showing only strict inequalities) identifies the games having only strict Nash equilibria. Note that, if $(a-b)=(b-d)$, then the linked game can be a prisoners' dilemma or a cooperation game, but it cannot be a coordination game or a chicken game; if $(a-b)>(b-d)$, then the linked game can be a prisoners' dilemma, a chicken game, or a cooperation game, but it cannot be a coordination game; and, finally, if $(b-d)>(a-b)$, then the linked game can be a prisoners' dilemma, a coordination game, or a cooperation game, but it cannot be a chicken game.

players would find it difficult to coordinate on the Pareto efficient Nash equilibrium. Nordhaus only shows the equilibria selected by a computer algorithm.

5. Linkage with retaliation permitted

As retaliation is a best response to the imposition of tariffs by the members of a coercive trade agreement, we can expect that, if non-members are permitted to retaliate, then they will retaliate. With retaliation, player i gets $\pi_i^s = \beta(k+1) + bk + c(N-k-1)$ by joining the CTA and $\pi_i^n = \alpha + \beta k + b(N-k-1) + ck$ by not joining it. Compared to when retaliation is prohibited, retaliation lowers the payoff to being a member of the CTA (for $k < N - 1$), and raises the payoff to being a non-member (for $k > 0$). Retaliation thus reduces the incentives to join the CTA.

Setting $\pi_i^s = \pi_i^n$ now gives

$$\hat{k} = \frac{(N-1)}{2} + \frac{(\alpha-\beta)}{2(b-c)}. \tag{2}$$

Focusing again on situations in which Nash equilibria are strict we have:

Proposition 2. *Depending on parameter values (see Table 2), if retaliation against members of a coercive trade agreement is permitted, the linked game is either: (i) a prisoners' dilemma in which membership in the CTA is zero; or (ii) a coordination game with tipping point $\hat{k}$. The latter game has two Nash equilibria in pure strategies, zero participation, which is risk-dominant, and full participation, which is payoff-dominant.*

Table 2. Properties of Linked Game: Retaliation Allowed

Condition	Game
$(\alpha-\beta) > (b-c)(N-1)$	Prisoners' dilemma
$(b-c)(N-1) > (\alpha-\beta)$	Coordination

In Table 2, $(\alpha-\beta)$ again represents the payoff to free riding in the supply of the global public good, and $(b-c)(N-1)$ represents the payoff to having free trade relations with all other states as opposed to engaging with them in a "trade war." As in our previous analysis, if the linked game is a prisoners' dilemma, linkage cannot aid cooperation in supplying the global public good, whereas if the linked game is a coordination game, linkage *may* help, but only if the players succeed in coordinating on the Pareto efficient Nash equilibrium.

It is easy to see from eq. (2) why the payoff-dominant Nash equilibrium cannot also be risk-dominant. From the perspective of any player, the tipping point for the coordination game is greater than one-half the number of other countries. Assuming that these other countries are as likely to join the agreement as not to join it, each country finds not joining to be the less risky option.

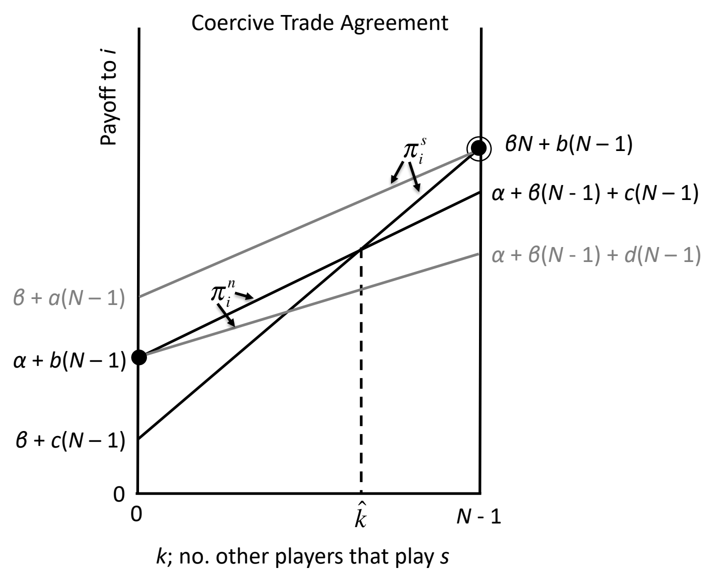
Comparing tables 1 and 2 we find (for a proof, see the Appendix):

Proposition 3. *Retaliation weakens but does not necessarily eliminate the incentives to form a coercive trade agreement. Specifically:*

- *i.) If the linked game is either a prisoners' dilemma or a chicken game when retaliation is prohibited, then it is a prisoners' dilemma when retaliation is permitted.*
- *ii.) If the linked game is a coordination game when retaliation is prohibited, then it is either a prisoners' dilemma or a coordination game when retaliation is permitted. When the linked game is a coordination game in both situations, the tipping point is higher when retaliation is permitted (that is, $\tilde{k}<\hat{k}$), making coordination on the efficient equilibrium more difficult.*
- *iii.) If the linked game is a cooperation game when retaliation is prohibited, then it is either a prisoners' dilemma or a coordination game (in which the payoff and risk dominant Nash equilibria differ) when retaliation is permitted.*

Figure 3 provides an illustration of this proposition. Participation in the CTA is a cooperation game when retaliation is prohibited (gray lines) but a coordination game with a fairly high tipping point when retaliation is allowed (black lines).

Fig. 3



To sum up, our theory shows that, even with retaliation, for a limited set of parameter values, linkage may succeed in sustaining efficient provision of the global public good without trade being restricted. This is the opportunity created by linkage. The risk is that, in attempting to link, countries may fail to coordinate successfully.

6. Institutions for linkage

The analysis above presumes that, when choosing whether or not to link, countries act independently—meaning, unilaterally. To explore choice under a multilateral regime, assume that the conditions for coordination are favorable—that is, $(b-c)(N-1)>(\alpha-\beta)$—and consider the following game: in stage 1, every country chooses independently whether or not to join the linked agreement; in stage 2, signatories choose collectively whether or not to supply the public good and impose tariffs; and, in stage 3, non-signatories choose independently whether or not to supply the public good and retaliate.[14] Solving the game backwards, it is easy to see that, in stage 3, non-signatories will supply the public good and thus avoid the imposition of tariffs, provided $k > \hat{k}$ (otherwise, they will not supply the public good and retaliate against the countries that impose tariffs upon them); in stage 2, signatories, knowing how non-signatories will "respond" in stage 3, will choose to supply the public good and impose tariffs on countries that do not supply it, provided that the number of signatories is at least $\hat{k}$ (otherwise, they will not supply the public good or impose tariffs on non-suppliers); and, finally, in stage 1, all countries will choose to participate (participation being a weakly dominant strategy). Intuitively, the linked agreement requires that signatories to the agreement supply the public good and impose tariffs on non-suppliers, with the agreement entering into force—and thus being legally binding on its parties—if and only if at least $\hat{k}$ countries have signed the agreement. Under this arrangement, a country can't lose by signing, for if the agreement fails to enter into force, signatories are under no obligation to supply the public good. However, a country can be made better off by signing, for if the agreement enters into force, all countries will want to sign it and all will then be better off compared to a situation in which the agreement does not enter into force. The multilateral linked agreement assures that coordination succeeds.

As unilateralism leaves coordination in doubt, multilateralism might seem the superior institution. However, multilateralism's superiority depends on whether at

[14] Though this formulation essentially assumes that signatories are committed to fulfilling their stage 2 obligations, commitment in this context is both different from and weaker than Nordhaus's assumption that all countries commit not to retaliate against a "climate club" for imposing tariffs. Our assumption is different because international law *requires* that countries abide by the treaties they sign and ratify, so long as these agreements enter into force. As stated in the Vienna Convention on the Law of Treaties, "every treaty in force is binding upon the parties to it and must be performed by them in good faith." Our assumption is weaker because countries exercise their sovereignty (in stage 1) when they choose whether or not to participate in an agreement. Only countries that choose to participate in a treaty are obligated to fulfill the obligations expressed in the agreement. By contrasting the stage game with a repeated game model, Barrett (2003) shows that enforcement of participation (stage 1) is the binding constraint on cooperation, not enforcement of compliance (stage 2).

least $\hat{k}$ players see the strategic advantage in choosing to participate in the linked agreement. Our experiment tests which institution is best for those limited circumstances in which linkage is potentially helpful for supplying the global public good (that is, situations in which $(b-c)(N-1)>(\alpha-\beta)$).

7. Experimental design

Our experiment sets N = 5, α = 5, β = 2, and c = 0. To test for the sensitivity of behavior to payoffs, we vary the value of the remaining parameter, b. In the *High* treatments, we set b = 3. In the *Low* treatments, we set b = 1. It is easy to confirm that our parameter choices imply that the linked game is a coordination game with a relatively low tipping point when the value for b is high and a relatively high tipping point when the value for b is low. The players in our experiment must choose between two situations: to keep their interactions on both trade and the public good separate or to link them by making cooperation on trade conditional on cooperation in supplying the public good.

To simplify the experiment, we assume that if cooperation on trade and the public good are unlinked, the players can take cooperation on trade as given, and thus be sure of getting the "free trade" payoff (4 in *Low* and 12 in *High*) plus whatever they get playing the public good game, a prisoners' dilemma. We call this the A game. When cooperation on trade is made conditional on supply of the public good, we assume that members of the linked agreement supply the public good and that non-members do not supply it. We also assume that members of the linked agreement impose positive tariffs on non-members only, and that non-members impose positive tariffs on members only, in retaliation. We call this game the B game.

Figures 4 and 5 show how the trade and public good cooperation games are used to calculate the A and B games. Note that, in our experiment, the players do not see the separate trade and public good games. Nor do the words "trade" and "public good" appear in any communications with our subjects (experimental instructions are shown in the Online Appendix). Once a game is chosen, the players must make a simultaneous choice. Each player is given two playing cards, one red and one black, and must decide which card to hand in. Handing in the red card means supplying the public good. Handing in the black card means not supplying the public good. The cells in the games shown in Figures 4 and 5 indicate the payoffs a player gets depending on which card he or she hands in and which cards his or her co-players hand in.

Fig. 4

Treatment *Low*

Trade	Number of red cards handed in by your co-players				
	0	**1**	**2**	**3**	**4**
Red	0	1	2	3	4
Black	0	0	0	0	0

Climate	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

A	Number of red cards handed in by your co-players				
	0	**1**	**2**	**3**	**4**
Red	6	8	10	12	14
Black	9	11	13	15	17

or

B	Number of red cards handed in by your co-players				
	0	**1**	**2**	**3**	**4**
Red	2	5	8	11	14
Black	9	10	11	12	13

Fig. 5

Treatment *High*

Trade	Number of red cards handed in by your co-players				
	0	**1**	**2**	**3**	**4**
Red	0	3	6	9	12
Black	0	0	0	0	0

Climate	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

A	Number of red cards handed in by your co-players				
	0	**1**	**2**	**3**	**4**
Red	14	16	18	20	22
Black	17	19	21	23	25

or

B	Number of red cards handed in by your co-players				
	0	**1**	**2**	**3**	**4**
Red	2	7	12	17	22
Black	17	16	15	14	13

How do the players decide which game to play? In the *Unilateral* treatment, the players play A, the default, so long as no player votes for B. However, if at least one player votes for B, all of the players must play B. We interpret this treatment as reflecting "bottom up," individual behavior. In the *Multilateral* treatment, the players play A, the default, unless a majority votes for B, where the size of the majority is determined by the tipping point for the corresponding B game (three out of five for *High*; four out of five for *Low*). Moreover, we assume that if B is chosen then the B-voters *must play Red* in the contribution rounds that follow, whereas the A-voters are free to play Red or Black when B is chosen. An A-voter who plays Red can be interpreted as a "cooperating non-member," meaning a player who abides by an agreement without being legally bound by it, whereas an A-voter who plays Black can be interpreted as a free rider.

As there are two differences between *Unilateral* and *Multilateral* (the majority vote and the requirement that B-voters play Red when B is chosen), we also consider a third treatment, *Majority*. In *Majority*, the players play A, the default, so long as a majority of players votes for A (where the size of the majority is again calculated with reference to the tipping point for the corresponding B game); otherwise the players play B.[15] In both *Unilateral* and *Majority*, every player is free to play Red or Black in every contribution round, irrespective of the game being played; the only difference between these treatments is the vote. Similarly, in both *Multilateral* and *Majority,* a majority determines which game is played; the only difference between *these* treatments is that the B-voters must play Red when B gets a majority in *Multilateral* whereas the B-voters are free to play Red or Black as they please in *Majority*. As our main interest lies in the *Unilateral* and *Multilateral* treatments, we confine our discussion in the paper to these treatments. We present our results for the intermediate treatment in the Online Appendix.

Table 3 summarizes the four main treatments as they were played in our experiment.

Table 3. Overview of main treatments

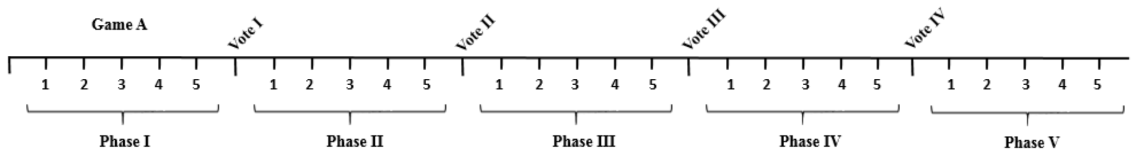
Treatment	Payoff	Group size	Tipping point	Voting rule	No. of groups	No. players
Unilateral-High	High	5	3	1/5	20	100
Multilateral-High	High	5	3	3/5 with commitment	20	100
Unilateral-Low	Low	5	4	1/5	20	100
Multilateral-Low	Low	5	4	4/5 with commitment	20	100

In every treatment, the experiment is played over five phases; see Figure 6. In phase I, groups play game A, the default, in five successive contribution rounds. In the first contribution round, the players choose Red or Black simultaneously, and then see how everyone chose to play. In the second and subsequent rounds, they repeat this cycle of play-and-see. In phases II-V, the players first vote for which game they want to play, A or B, and then, after the results of the vote have been made public, they play five successive contribution rounds, just as in Phase I. As noted before, in *Multilateral*,

[15] In the context of climate change, *Majority* would apply if the use of tariffs were adopted in a "decision" by the conference of the parties to the UN Framework Convention on Climate Change. Under this agreement, decisions are to be made by consensus if possible but by a three-quarters vote if necessary, and are legally non-binding, meaning that a vote to link is not a commitment to link.

if B is selected, the players that voted for B must play Red in all five rounds. Only after this can a player decide to withdraw from the agreement (meaning, vote for A), a rule of the game that accords with international law.[16] In summary, in our experiment the players choose which game to play four times and which card to hand in 25 times.

Figure 6. Timeline



The experimental sessions were held in a computer lab at two universities in Germany, using undergraduate students recruited from the general student population. In total, 600 students participated in the experiment, each student taking part in one treatment only. There were six treatments (the four main treatments shown in Table 3 and the two additional *Majority* treatments shown in the Online Appendix) 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. 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 subjects were seated at linked computers (game software Ztree; Fischbacher 2007) and randomly assigned to one of five 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

[16] Multilateral treaties normally incorporate an explicit provision for withdrawal, with the default being that a state must give at least one year's notice of an intention to withdraw. On 1 June 2017, President Trump announced that the US would withdraw from the Paris Agreement, but according to Article 28, the earliest any party's withdrawal can take effect is 4 November 2020, one day after the next US presidential election.

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 25 contribution rounds and applying an exchange rate of €.04 per token. After the game, the subjects were asked to complete a questionnaire before they were paid their earnings in cash.

8. Experimental results

Figures 7-10 show how individuals and groups behaved in the four main treatments. In every figure, the horizontal axis shows contribution rounds for the five phases and the vertical axis (on the right) shows payoffs. The data points show average payoffs for each group in each round. The data points in blue indicate that the group played the A game in this phase. The data points in orange indicate that the group played the B game. Finally, the green bars show the number of group members who voted for game B in each phase (left axis). Recall that all groups were required to play game A in phase I, but were free to choose between A and B in phases II-V.

In both *Unilateral-High* (Figure 7) and *Multilateral-High* (Figure 8) all groups but one played B in at least one phase. Of the 19 groups that played B in *Unilateral-High*, 15 coordinated very successfully and four failed to coordinate. Of the 19 groups that played B in *Multilateral-High*, almost all coordinated very well, as could be expected from the design. In *Unilateral-Low* (Figure 9), all groups tried B, but less than half coordinated successfully. In *Multilateral-Low* (Figure 10), three-fourths of the groups tried B and coordinated successfully.[17]

[17] It is perhaps no surprise that coordination should prove more difficult in *Unilateral-Low* than in *Unilateral-High*, given that the tipping point is higher in the former treatment than in the latter one. But why then don't many fewer groups attempt to play B in *Unilateral-Low* than in *Unilateral-High*? The reason may be that only a single player must choose to play B in order for B to be played, but another reason may be that the penalty for failing to coordinate in B is much lower in *Unilateral-Low* than in *Unilateral-High*.

Figure 7. Payoffs over time by group for *Unilateral-High*

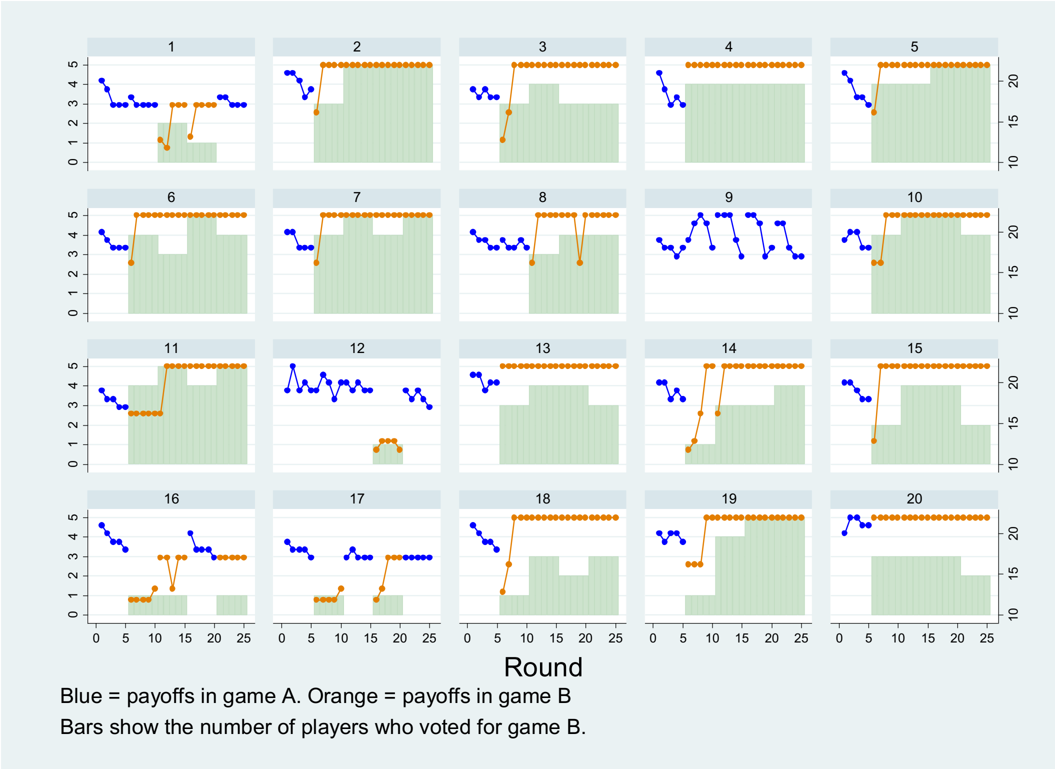


Figure 8. Payoffs over time by group for *Multilateral-High*

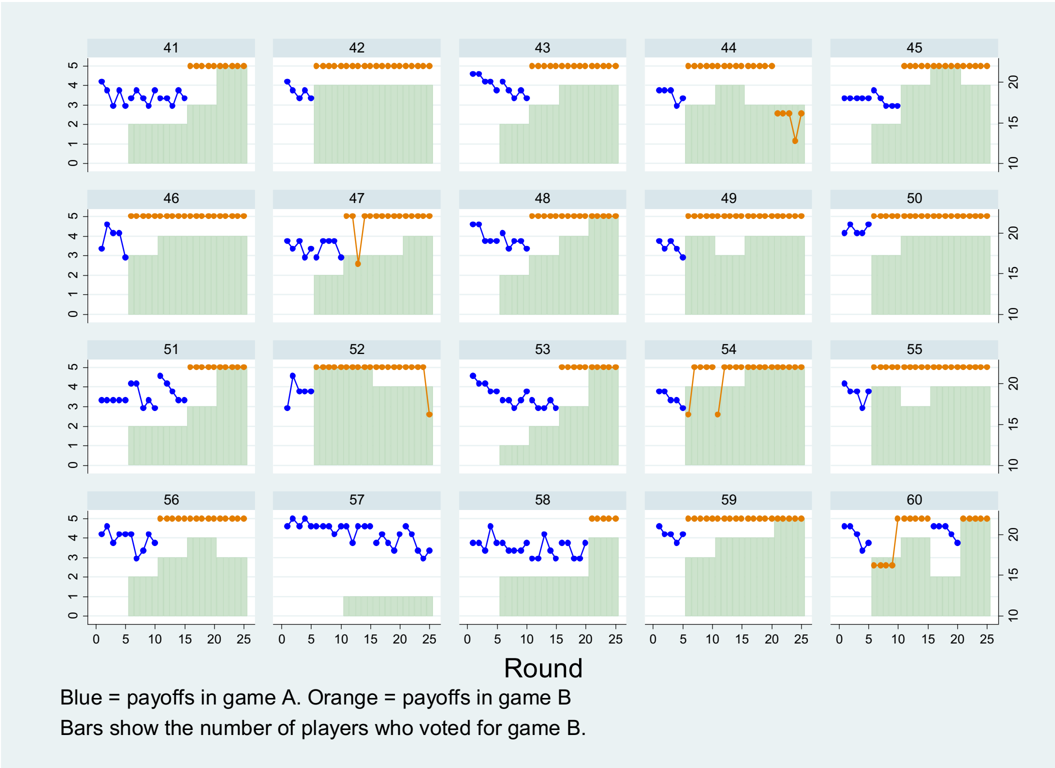


Figure 9. Payoffs over time by group for *Unilateral-Low*

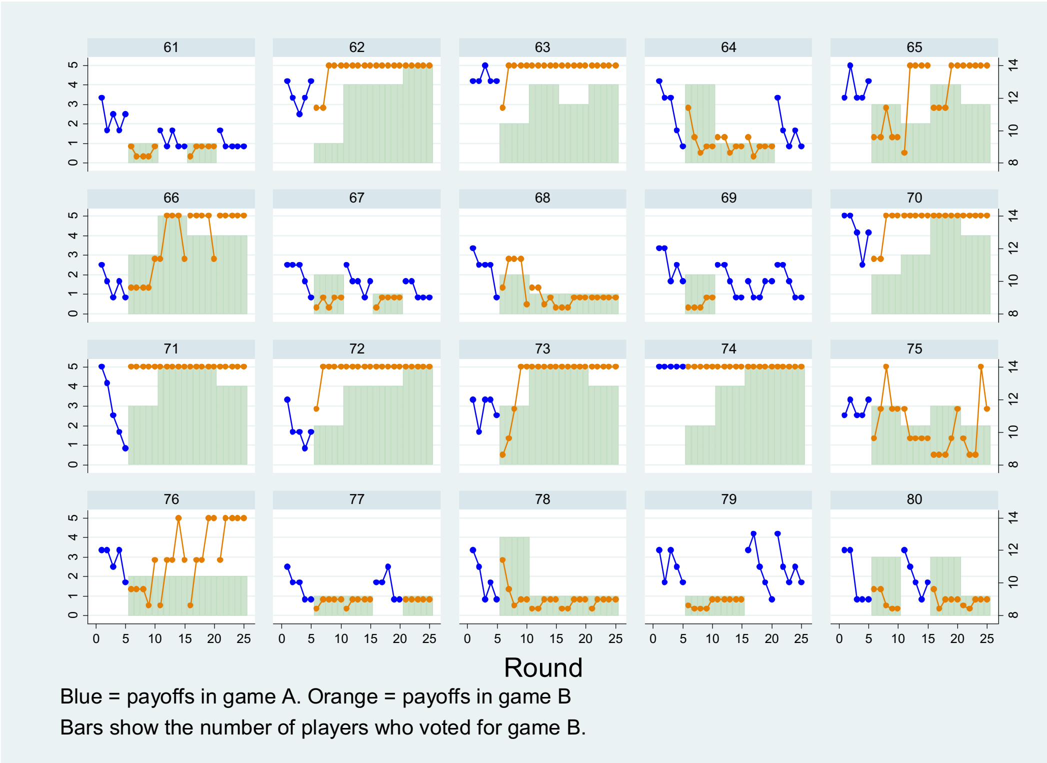
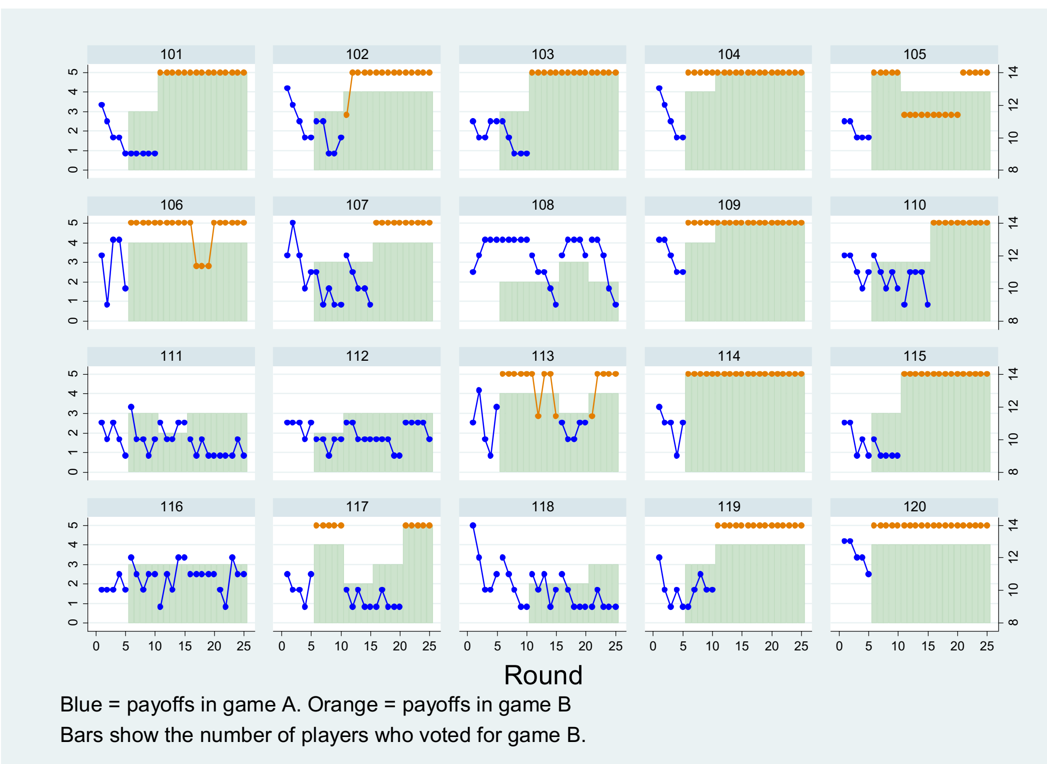


Figure 10. Payoffs over time by group for *Multilateral-Low*



For both the *High* and *Low* treatments, the results show that linkage was more likely to be tried *initially* when the players could decide whether to link unilaterally. Over time, however, support for B wavered in the *Unilateral* treatments whereas it grew in the *Multilateral* treatments. By the end of the game, at least as many groups played B in *Multilateral* as in *Unilateral*. Fisher's exact tests show that, in phase II (the first phase in which players choose which game to play), the proportion of groups that played game B is significantly higher in *Unilateral* than in *Multilateral*, whether payoffs were *High* (P = 0.096) or *Low* (P = 0.000), whereas there are no significant differences between *Unilateral* and *Multilateral* in either the *High* or *Low* treatments in the later phases (III-V).[18]

Our results thus indicate a trade-off, especially in the *Low* treatments. A regime of unilateralism makes it very likely that linkage will be attempted, and that this will happen early in the game, but unilateralism makes linkage risky, as coordination often fails. A regime of multilateralism makes it very likely that coordination will succeed when linkage is attempted, but multilateralism reduces the chances that linkage will be attempted, especially early in the game.

Importantly, the failure to coordinate in game B in *Unilateral* and the failure to choose B in *Multilateral* early in the game have very different consequences. When groups in the *Multilateral* treatments failed to try B in the early phases, there was still a good chance that they would try B later on. Only when they played A in all of the first four phases were they unlikely to switch to B in the next and last phase. By contrast, groups in the *Unilateral* treatments rarely recovered from a failure to coordinate in game B. In 15 out of 18 cases of coordination failure, groups either continued to choose B and failed or they gave up on B and switched back to A.

Table 4 provides an overview by showing the proportion of groups that played games A and B, and these groups' average contributions and payoffs, in each phase for all

[18] Unless stated otherwise, all statistical tests are two-sided.

four treatments. In the *High* treatments (*Unilateral-High* and *Multilateral-High*), contributions are significantly greater in all phases for the groups that played B than for the groups that played A. Payoffs are also greater for the groups that played B in all phases, except for phase II in *Unilateral-High*. (P-values are shown in Table 4.) Similarly, in *Multilateral-Low*, contributions and payoffs are significantly greater for the groups that played B than for the groups that played A. However, in *Unilateral-Low* contributions are significantly greater for the B-groups than for the A-groups only in phases III and V and payoffs are *never* significantly different between the B-groups and A-groups. For *Multilateral*, the advantage of B over A is robust to whether payoffs are *High* or *Low*. For *Unilateral*, this is not the case. In the Online Appendix (see Table A.5), we show that when groups play B in the *Unilateral* treatments, contributions are significantly greater when payoffs are *High* than when they are *Low* ($P < 0.05$ in phases II, IV, V and $P < 0.1$ in phase III).[19] Essentially, coordination on the welfare-superior Nash equilibrium in the B game is easier with a low tipping point in *High* than with a high tipping point in *Low*. This result parallels a finding in the literature on threshold public goods games which shows that groups are more likely to reach the threshold when the threshold is lower (Croson and Marks, 2000).

[19] The same table shows that there is no statistically significant difference in contributions between the *High* and *Low* treatments for either institution when groups play A or when they play B in *Multilateral*.

Table 4. Contributions and payoffs by treatment

		Unilateral-High				*Multilateral-High*		
Phase	Game	% of groups	Average % contributions	Average payoff		% of groups	Average % contributions	Average payoff
I	A	100	40	19.0		100	44	19.2
	B	0	-	-		0	-	-
II	A	20	37	18.9		50	32	18.6
	B	80	85***	18.9		50	98***	21.4***
III	A	15	40	19		25	32	18.6
	B	85	89**	21.1**		75	99***	21.8***
IV	A	10	42	19.1		15	45	19.3
	B	90	87**	20.7*		85	100***	22.0***
V	A	20	17	17.9		5	36	18.8
	B	80	94***	21.7***		95	99***	21.6**
		Unilateral-Low				*Multilateral-Low*		
Phase	Game	% of groups	Average % contributions	Average payoff		% of groups	Average % contributions	Average payoff
I	A	100	46	11.3		100	40	11.0
	B	0	-	-		0	-	-
II	A	0	-	-		60	24	10.2
	B	100	58	10.8		40	100***	14***
III	A	20	19	10.0		40	27	10.3
	B	80	69**	11.9		60	97***	13.7***
IV	A	15	23	10.1		35	26	10.3
	B	85	63	11.7		65	98***	13.7***
V	A	25	18	9.9		25	25	10.2
	B	75	72*	12.4		75	100***	14***

Note: Stars indicate statistically significant differences between groups that play A and groups that play B within the same treatment and same phase according to a two-sided Mann-Whitney-Wilcoxon test. Significance levels: * P < 0.1, ** P < 0.05, *** P < 0.01.

Countries cannot choose the parameters of the game (*High* versus *Low*), but they can choose the institutions (*Unilateral* versus *Multilateral*). Which institution performs best? Comparing average payoffs in each phase (see Tables A.3 and A.4 in the Online Appendix), we find that *Multilateral* performs better than *Unilateral* when payoffs are *Low* whereas there is no significant difference when payoffs are *High*. Given *Multilateral*'s advantage in assuring coordination, it may seem surprising that *Multilateral* does not have an overall payoff advantage in the *High* treatments. The reason is that, although coordination fails one-fifth of the time in *Unilateral*, it takes longer for groups to try B in *Multilateral*, and a mixed record of coordination in B compares favorably to a failure to cooperate in A. We also find that the quartile of groups that receive the lowest payoff in each treatment do significantly better under *Multilateral* than *Unilateral* in both the *High* and *Low* treatments (P < 0.05). The reason for this is that the worst-performing groups in *Multilateral* either do not try B or they try it late in the game whereas the low-performing groups in *Unilateral* try game B but fail to coordinate successfully. *Multilateral* thus not only does at least as well as *Unilateral* overall, but also insures against a bad outcome.

Why is B not chosen more often in *Multilateral*? From a theoretical perspective, a player can't lose and may gain by voting for B in *Multilateral*. But in both the *High* and *Low* treatments, should the players have "naïve" expectations, game A can appear more attractive than B. If, for example, a player assumes that her co-players in each possible situation toss a coin to decide whether to play Red or Black, then voting for A and playing Black in A gives a higher expected payoff than voting for B and playing Red in B. In an ex-post questionnaire, we asked our participants whether they expected their co-players to play Red in game A and in game B. Focusing on the players in the *Multilateral* treatments who never played B, we find no significant differences in their beliefs about how their co-players would play in game A and game B (proportion test, $P > 0.1$). By contrast, the players who played game B at least once over the course of the game reported having different beliefs about how their co-players would play in games A and B ($P < 0.001$). Although these results are based on ex-post responses and do not imply causality, they are in line with previous findings about the crucial role of beliefs in determining how groups choose between playing different games (Barrett and Dannenberg, 2017; Dal Bó et al., 2018).[20]

The final question is whether the differences we observe in the *Unilateral* and *Multilateral* treatments (especially with *Low* payoffs) are caused by the majority-voting rule or the requirement that B-voters play Red should B be chosen. Results for our *Majority-High* and *Majority-Low* treatments (presented in detail in the Online Appendix) show that the majority rule by itself does not perform better than *Unilateral*, irrespective of whether payoffs are *High* or *Low* (MWW test, $P > 0.1$ in

[20] For a review of experimental studies on endogenous institutional choice, see Dannenberg and Gallier (2019).

every phase). By contrast, *Majority* performs worse than *Multilateral* and the differences are significant in one phase with *High* payoffs ($P < 0.05$) and in two phases with *Low* payoffs ($P < 0.1$). The reason for this is that, in *Majority*, fewer groups (relative to *Multilateral*) tried game B and fewer of these groups coordinated successfully, especially in the *Low* treatment for which the tipping point is high.[21]

9. Conclusions

As trade is bilateral, trade agreements can be sustained by a strategy of direct reciprocity. Cooperation to supply a global public good, by contrast, must be sustained by a strategy of diffused reciprocity, and is therefore harder. Intuitively, linkage of these different issues—making cooperation on trade contingent on cooperation in supplying a global public good—presents both an opportunity and a risk. Linkage could fortify efforts to supply a global public good or it could disrupt cooperation on trade.

We show here that, for linkage to be a rational strategy, it must be the case that the gains from cooperating on trade are large relative to the gains from cooperating to supply a global public good, making the linked game a coordination game. Whether this condition holds empirically for an issue like climate change is a matter requiring further research.

When the linked game is a coordination game, theory suggests that coordination is only assured when decisions are made multilaterally. Our experiment largely confirms the advantage of multilateralism over unilateralism, though we also find that each regime has advantages and disadvantages. Under unilateralism, linkage is more likely to be attempted, especially early in the game. Under multilateralism, coordination is more likely to succeed when linkage is attempted. On average,

[21] This last result implies that what is important about multilateral agreements is not that they allow countries to express their preferences or expectations for how others will behave but that they provide assurance for how parties will behave should the agreement enter into force.

multilateralism performs best. Moreover, multilateralism insures against a particularly bad outcome in which linkage is attempted but coordination fails. The risk of this failure is particularly large when the gains to cooperation on trade are low relative to the gains to cooperation in supplying the global public good. In our *Low* treatment, when the decision to link was made unilaterally, coordination failed half the time. We have resisted calling coordination failure a "trade war," because in our experiment we assumed, rather than demonstrated, that players would cooperate on trade in the unlinked situation. However, a trade war is clearly a possible implication of coordination failure, one that should be explored in future work.[22]

Two other aspects of our modeling approach deserve closer examination, our assumption that countries are symmetric and the lack of an explicit model of trade. Allowing countries to be asymmetric would affect the gains from trade cooperation, the gains from cooperating to supply the global public good, and the tipping point for coordination. Incorporating an explicit model of trade would allow limits on emissions by a subset of countries to cause "trade leakage," meaning a shift in comparative advantage that in turn causes emissions by non-members to change (most likely, to increase).[23] Leakage would likely weaken the incentive for a subset of countries to cooperate on climate change in the absence of linkage but it would at the same time likely strengthen the incentive these countries have to link cooperation on trade to cooperation on climate change.[24]

[22] For a recent empirical analysis of the costs of a trade war, see Amiti, Redding, and Weinstein (2019). For an analysis of the costs of trade policy uncertainty created by the threat of a trade war, see Handley and Limão (2017).
[23] Aichele and Felbrermayr (2015) find that the Kyoto Protocol, which limited the emissions of some but not all countries, did cause leakage.
[24] See Barrett (2003), Chapter 12.

Appendix

Proof of Proposition 3. *The only result that requires demonstration is the claim that* $\tilde{k} < \hat{k}$. *To prove this claim, suppose to the contrary that* $\tilde{k} \geq \hat{k}$. *Then we have*

$$\frac{(\alpha-\beta)-(a-b)(N-1)}{(b-d)-(a-b)} > \frac{(\alpha-\beta)+(b-c)(N-1)}{2(b-c)}. \tag{A.1}$$

Rearranging gives

$$\frac{\left[(a-c)-(c-d)\right]}{(a-d)} \geq \frac{(b-c)(N-1)}{(\alpha-\beta)}. \tag{A.2}$$

From Table 2 we know that the RHS of (A.2) is strictly greater than one. This means that the LHS of (A.2) must be greater than one, or that $(c-d)<0$, *which is false by assumption. Hence, it must be the case that* $\hat{k} > \tilde{k}$.■

References

Aichele, Rahel and Gabriel Felbermayr (2015), "Kyoto and Carbon Leakage: An Empirical Analysis of the Carbon Content of Bilateral Trade," *Review of Economics and Statistics* 97(1): 104-115.

Amiti, Mary, Stephen J. Redding, and David E. Weinstein (2019), "The Impact of the 2018 Trade War on U.S. Prices and Welfare," CEPR Discussion Paper DP13564, March.

Bagwell, Kyle and Robert W. Staiger (1999), "An Economic Theory of GATT," *American Economic Review* 89(1): 215-248.

Barrett, Scott (1994), "Self-Enforcing International Environmental Agreements," *Oxford Economic Papers*, 46: 878-894.

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: Oxford University Press.

Barrett, Scott and Astrid Dannenberg (2016), "An Experimental Investigation into 'Pledge and Review' in Climate Negotiations," *Climatic Change* 138(1): 339-351.

Barrett, Scott and Astrid Dannenberg (2017), "Tipping versus Cooperating to Supply a Public Good," *Journal of the European Economics Association* 15(4): 910-941.

Bhagwati, Jagdish, Pravin Krishna, and Arvind Panagariya, eds. (1999), *Trading Blocks: Alternative Approaches to Analyzing Preferential Trade Agreements*, Cambridge MA: MIT Press.

Böhringer, Christoph, Jared C. Carbone, and Thomas F. Rutherford (2016), "The Strategic Value of Carbon Tariffs," *American Economic Journal: Economic Policy*, 8(1): 28-51.

Böhringer, Christoph and Thomas F. Rutherford (2016), "Paris after Trump: An Inconvenient Insight," Working Paper V-400-17, Department of Economics, University of Oldenburg.

Crawford, Vincent P. (2016), "New Directions for Modelling Strategic Behavior: Game-Theoretic Models of Communication, Coordination, and Cooperation in Economic Relationships," *Journal of Economic Perspectives* 30(4): 131-150.

Dannenberg, Astrid and Carlo Gallier (2019), On the Endogenous Choice of Institutions: A Survey of Experimental Research, Working paper.

Eichner, Thomas and Rüdiger Pethig (2013), "Self-Enforcing Environmental Agreements and International Trade," *Journal of Public Economics* 102: 37-50.

Finus, Michael (2001), *Game Theory and International Environmental Cooperation*, Cheltenham, UK: Edward Elgar.

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

Handley, Kyle and Nuno Limão (2017), "Policy Uncertainty, Trade and Welfare: Theory and Evidence for China and the U.S.," *American Economic Review* 107(9): 2731-2783.

Krueger, Anne O. (1999), "Are Preferential Trading Arrangements Trade-Liberalizing or Protectionist?" *Journal of Economic Perspectives* 13(4): 105-124.

Ledyard, John O. (1995), "Public Goods: A Survey of Experimental Research," in John Kagel and Alvin Roth (eds.), *Handbook of Experimental Economics*, Princeton: Princeton University Press, pp. 111-194.

Lessmann, Kai, Robert Marschinski, and Ottmar Edenhofer (2009), "The Effects of Tariffs on Coalition Formation in a Dynamic Global Warming Game, *Economic Modelling* 26: 641-649.

Maggi, Giovanni (2016), "Issue Linkage," in Kyle Bagwell and Robert W. Staiger (eds.), *The Handbook of Commercial Policy*, vol. 1, Amsterdam: North Holland, 513-564.

Nordhaus, William (2015), "Climate Clubs: Overcoming Free-Riding in International Climate Policy," *American Economic Review* 105(4): 1339-1370.

Ossa, Ralph (2011), "A 'New Trade' Theory of GATT/WTO Negotiations," *Journal of Political Economy* 119(1): 122-152.

Ossa, Ralph (2014), "Trade Wars and Trade Talks with Data," *American Economic Review* **104**(2): 4104-4146.

Samuelson, Larry (2016), "Game Theory in Economics and Beyond," *Journal of Economic Perspectives* 30(4): 107-130.

Schelling, Thomas C. (1978), *Micromotives and Macrobehavior*, New York: WW Norton & Company.

Wagner, Ulrich J. (2015), "Estimating Strategic Models of International Treaty Formation," *Review of Economic Studies* 83: 1741-1778.

Winchester, Nevin (2018), "Can Tariffs be Used to Enforce Paris Climate Commitments?" *World Economy* 41: 2650-2668.

World Trade Organization (2009), *World Trade Report 2009: Trade Policy Commitments and Contingency Measures*, Geneva: World Trade Organization.

Appendix for Online Publication

Coercive Trade Agreements for Supplying Global Public Goods

Scott Barrett[*]
Columbia University

Astrid Dannenberg[+]
University of Kassel

This Appendix reproduces the English translations of the original experimental instructions presented to the subjects in the German language (only for *Multilateral-High*, the instructions for the other treatments are very similar) and provides additional analyses of the experimental data (Tables A1-A8 and Figures A1-A7).

1. Sample instructions, translated into English: *Multilateral-High* treatment

Experimental Instructions

Welcome to our experiment!

1. General information

In our experiment you can earn money. How much money you earn will depend on the gameplay, 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 not talk to other participants. Now, read the following rules of the game carefully. If you have any questions, raise your hand and we will come to you and answer your questions.

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**. In each game, you will receive two cards, a **Red card** and a **Black card**, and you will 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.

Voting

A Game		Number of Red cards handed in by your co-players				
		0	1	2	3	4
Your choice	Red	14	16	18	20	22
	Black	17	19	21	23	25

B Game		Number of Red cards handed in by your co-players				
		0	1	2	3	4
Your choice	Red	-	-	12	17	22
	Black	-	-	-	14	13

Here is an example for how to read the tables:

If the group plays the A Game and two of your co-players hand in their Red cards (and the other two co-players hand in their Black cards), you will get 18 tokens if you hand in your Red card and you will get 21 tokens if you hand in your Black card.

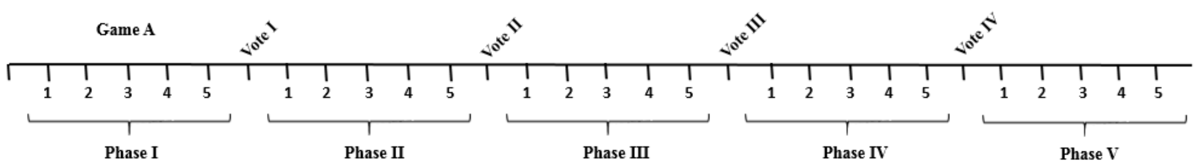
Similarly, if the group plays the B Game and three of your co-players hand in their Red cards, you will get 17 tokens if you hand in your Red card and you will get 14 players if you hand in your Black card. A difference between the A and B games is that some cells in the table for the B Game are blank. The reason for this is explained below.

You and your co-players will first play Game A five times consecutively. After that, every player in your group will vote on whether to play Game A or Game B. All the players' votes will be displayed to everyone before the game starts. The game that receives the **most votes** (at least 3 out of 5) will then be played five times consecutively. If Game A is chosen, all players will be free to play the Red card or the Black card, irrespective of whether they voted for Game A or Game B. However, if Game B is chosen, the players who voted for B **must** play the Red card whereas the players who voted for A will be free to play the Red card or the Black card. This explains why some of the cells in the B Game table are blank. Game B is only played if at least three players vote for B, in which case all the B-voters must play the Red card. This holds for all five rounds.

Because B-voters have to play Red when Game B is played, the following conditions hold for Game B:

- There are at most two A-voters who are free to choose between Red and Black.
- If there is only one A-voter, she will get 22 tokens if she plays Red and she will get 13 tokens if she plays Black. (The four B-voters will get 22 tokens if the A-voter plays Red and they will get 17 tokens if the A-voter plays Black.)
- If there are two A-voters and they both play Red then they will get 22 tokens each. (The three B-voters in this case will also get 22 tokens each.) If one A-voter plays Red and the other plays Black, the A-voter who plays Red will get 17 tokens and the A-voter who plays Black will get 13 tokens. (The three B-voters in this case will get 17 tokens each.) If both A-voters play Black, they will get 14 tokens each. (The three B-voters in this case will get 12 tokens each.)

The process of voting between Game A and Game B and then playing the chosen game will be repeated three more times. In total, your group will vote four times and play the chosen game five times after each vote. Hence, overall, there will be 25 rounds, as shown in the timeline below.



You will play with the same group of players throughout all rounds. The sum of tokens you earn across all 25 rounds will be paid to you in cash at the end. You will get €1 for 25 tokens. For example, if you earn 500 tokens in total, you will get €20.

3. *Control questions*

Please answer the following control questions.

a. Right or wrong? You and your co-players will first play Game A five times consecutively.

○ Right ○ Wrong

b. Right or wrong? After first playing Game A five times consecutively, all players will vote on whether to play Game A or Game B. The game that receives the most votes will then be played five times consecutively. Overall, your group will vote for Game A or Game B four times. After each vote, your group will play the chosen game five times consecutively.

○ Right ○ Wrong

c. Right or wrong? If Game A has been chosen, all players will be free to play the Red card or the Black card. If Game B has been chosen, the players who voted for B must play the Red card whereas the players who voted for A will be free to play the Red card or the Black card.

○ Right ○ Wrong

d. Assume that, when playing a round of the A Game, one of your co-players hands in his/her Red card (the other three co-players hand in their Black cards). 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, when playing a round of the A Game, three of your co-players hand in their Red cards (the other co-player hands in his/her 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, when playing a round of the B Game, three of your co-players hand in their Red cards (the other co-player hands in his/her 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, when playing a round of the A Game, all four of your co-players hand in their Red cards (none of your co-players 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, when playing a round of the B Game, all four of your co-players hand in their Red cards (none of your co-players 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. The game will begin after we have checked the answers of all the participants and answered any questions you may have. Good luck!

2. Further analyses of the experimental data

Table A1 summarizes our six treatments. With five people per group, each treatment is played by 100 players. The term "with commitment" means that, if a majority of players in any phase votes for B, then all of the B-voters in that phase must play Red. Of course, at the end of phases II-IV, all players get to vote again; the "commitment" is short-lived.

Table A2 summarizes the main results. We discuss these below.

Table A1. Overview of all treatments

Treatment	Payoff	Tipping point	Voting rule	Number of groups
Unilateral-High	High	3	1/5	20
Majority-High	High	3	3/5	20
Multilateral-High	High	3	3/5 with commitment	20
Unilateral-Low	Low	4	1/5	20
Majority-Low	Low	4	4/5	20
Multilateral-Low	Low	4	4/5 with commitment	20

Table A2. Contributions and payoffs by treatment

		Unilateral-High			*Majority-High*			*Multilateral-High*		
Phase	Game	% of groups	Average % contributions	Average payoff	% of groups	Average % contributions	Average payoff	% of groups	Average % contributions	Average payoff
I	A	100	40	19.0	100	44	19.2	100	44	19.2
	B	0	-	-	0	-	-	0	-	-
II	A	20	37	18.9	65	39	19.0	50	32	18.6
	B	80	85***	18.9	35	97***	21.2***	50	98***	21.4***
III	A	15	40	19	40	33	18.6	25	32	18.6
	B	85	89**	21.1**	60	99***	21.7***	75	99***	21.8***
IV	A	10	42	19.1	30	22	18.1	15	45	19.3
	B	90	87**	20.7*	70	99***	21.7***	85	100***	22.0***
V	A	20	17	17.9	15	23	18.1	5	36	18.8
	B	80	94***	21.7***	85	97***	21.4***	95	99***	21.6**
		Unilateral-Low			*Majority-Low*			*Multilateral-Low*		
Phase	Game	% of groups	Average % contributions	Average payoff	% of groups	Average % contributions	Average payoff	% of groups	Average % contributions	Average payoff
I	A	100	46	11.3	100	49	11.5	100	40	11.0
	B	0	-	-	0	-	-	0	-	-
II	A	0	-	-	75	29	10.5	60	24	10.2
	B	100	58	10.8	25	86***	12.4**	40	100***	14***
III	A	20	19	10.0	55	29	10.5	40	27	10.3
	B	80	69**	11.9	45	84***	12.8***	60	97***	13.7***
IV	A	15	23	10.1	55	29	10.5	35	26	10.3
	B	85	63	11.7	45	89***	13.0**	65	98***	13.7***
V	A	25	18	9.9	60	22	10.1	25	25	10.2
	B	75	72*	12.4	40	90***	13.2***	75	100***	14***

Note: The table shows the proportions of groups that player either game A or game B in each phase in the different treatments and provides comparisons of contributions and payoffs between groups that play A and groups that play B within the same treatment and same phase. Stars indicate statistically significant differences between groups that play A and groups that play B within the same treatment and same phase according to a two-sided Mann-Whitney-Wilcoxon test. Significance levels: * $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$.

Comparisons of how many groups choose game B between institutions

High payoffs: In *Unilateral-High*, the proportion of groups that play game B is high from the first to the last phase. In *Majority-High* and *Multilateral-High*, the proportion is low in the first phase but increases over time to 85% in *Majority-High* and 95% in *Multilateral-High* in the last phase. The proportion in *Multilateral-High* is always higher than in *Majority-High*. Fisher's exact tests show that, in phase II, the proportion of groups that play game B is significantly higher in *Unilateral-High* than in *Majority-High* (P = 0.010) and *Multilateral-High* (0.096). There are no significant differences in phase III, phase IV, and phase V.[25]

Low payoffs: In *Unilateral-Low*, all groups play game B in phase II and this proportion decreases slightly over time to 75% in the last phase. In *Majority-Low* and *Multilateral-Low*, the proportion is low in the first phase but increases over time to 40% in *Majority-Low* and 75% in *Multilateral-Low* in the last phase. The proportion in *Multilateral-Low* is always higher than in *Majority-Low*. Fisher's exact tests show that, in phase II, the proportion of groups that play game B is significantly higher in *Unilateral-Low* than in *Majority-Low* (P = 0.000) and *Multilateral-Low* (0.000). In phase III and phase IV, the proportion of B-groups is significantly higher in *Unilateral-Low* than in *Majority-Low* ($P < 0.05$ each). In the last phase V, the proportion of B-groups in both *Unilateral-Low* and *Multilateral-Low* are higher than in *Majority-Low* (P = 0.054 each).

Comparisons of contributions and payoffs between groups that play A and B

High payoffs: In all treatments, *Unilateral-High*, *Majority-High*, and *Multilateral-High*, and in all phases, contributions are significantly higher in the groups that play B than in the groups that play A. Payoffs are also higher in all treatments and all phases, except in phase II in *Unilateral-High*. (P-values are shown in Table 1.) That means that, on average, playing B pays off under all institutions.

Low payoffs: In *Majority-Low* and *Multilateral-Low*, contributions and payoffs are always significantly higher in the groups that play B than in the groups that play A. In *Unilateral-Low*, contributions are significantly higher in the B-groups than in the A-groups in phase III and phase V. Payoffs are never significantly different between B-groups and A-groups. That means that, on average, playing B pays off under *Majority-Low* and *Multilateral-Low* but not under *Unilateral-Low*.

Table A3 compares contributions, and Table A4 payoffs, between institutions, separately for the high payoff treatments and the low payoff treatments. We discuss these results below.

[25] All non-parametric tests are two-sided.

Table A3. Comparisons of contributions between institutions

	High			*Low*		
Phase	A	B	A & B	A	B	A & B
	Uni ≈ Maj	-	-	*Uni ≈ Maj*	-	-
I	*Uni ≈ Multi*	-	-	*Uni ≈ Multi*	-	-
	Maj ≈ Multi	-	-	*Maj ≈ Multi*	-	-
	Uni ≈ Maj	*Uni < Maj* **	*Uni ≈ Maj*	-	*Uni ≈ Maj*	*Uni ≈ Maj*
II	*Uni ≈ Multi*	*Uni < Multi* ***	*Uni ≈ Multi*	-	*Uni < Multi* ***	*Uni ≈ Multi*
	Maj ≈ Multi	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj < Multi* ***	*Maj ≈ Multi*
	Uni ≈ Maj	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*
III	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni < Multi* **	*Uni ≈ Multi*
	Maj ≈ Multi	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj < Multi* *
	Uni ≈ Maj	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*
IV	*Uni ≈ Multi*	*Uni < Multi* **	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni < Multi* **	*Uni ≈ Multi*
	Maj ≈ Multi	*Maj < Multi* **	*Maj< Multi***	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj ≈ Multi*
	Uni ≈ Maj	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*
V	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni < Multi* **	*Uni < Multi* *
	Maj ≈ Multi	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj < Multi* *	*Maj < Multi* **

Notes: ≈ indicates no significant difference, < or > indicate statistically significant difference according to a two-sided MWW test. Significance levels: * P < 0.1, ** P < 0.05, *** P < 0.01.

Table A4. Comparisons of payoffs between institutions

	High			*Low*		
Phase	A	B	A & B	A	B	A & B
	Uni ≈ Maj	-	-	*Uni ≈ Maj*	-	-
I	*Uni ≈ Multi*	-	-	*Uni ≈ Multi*	-	-
	Maj ≈ Multi	-	-	*Maj ≈ Multi*	-	-
	Uni ≈ Maj	*Uni < Maj* **	*Uni ≈ Maj*	-	*Uni ≈ Maj*	*Uni ≈ Maj*
II	*Uni ≈ Multi*	*Uni < Multi* ***	*Uni ≈ Multi*	-	*Uni < Multi* ***	*Uni < Multi* **
	Maj ≈ Multi	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj < Multi* ***	*Maj ≈ Multi*
	Uni ≈ Maj	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*
III	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni < Multi* **	*Uni < Multi* *
	Maj ≈ Multi	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj < Multi* *
	Uni ≈ Maj	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*
IV	*Uni ≈ Multi*	*Uni < Multi* **	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni < Multi* **	*Uni < Multi* *
	Maj ≈ Multi	*Maj < Multi* **	*Maj < Multi* **	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj ≈ Multi*
	Uni ≈ Maj	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*	*Uni ≈ Maj*
V	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni ≈ Multi*	*Uni < Multi* **	*Uni < Multi* *
	Maj ≈ Multi	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj ≈ Multi*	*Maj < Multi* **	*Maj < Multi* ***

Notes: ≈ indicates no significant difference, < or > indicate statistically significant difference according to a two-sided MWW test. Significance levels: * P < 0.1, ** P < 0.05, *** P < 0.01.

High payoffs: Contributions and payoffs never significantly differ between *Unilateral-High*, *Majority-High*, and *Multilateral-High* when groups play game A. When groups play game B, *Unilateral-High* leads to lower contributions and payoffs than the other two institutions in phase II and *Multilateral-High* leads to higher contributions and payoffs than the other two institutions in phase IV. Over both games (A&B), however, the differences in contributions and payoffs are rarely significant. The reason for this is that groups do relatively well under all three institutions. The B-groups in *Unilateral-High* have lower contributions and payoffs than the B-groups in *Multilateral-High* in phase II and phase IV but more groups play game B in *Unilateral-High* in those phases.

Low payoffs: Contributions and payoffs never significantly differ between *Unilateral-Low*, *Majority-Low*, and *Multilateral-Low* when groups play game A. When groups play game B, in each phase, *Multilateral-Low* leads to higher contributions and payoffs than at least one of the other institutions. These differences also show when we compare the payoffs over both games. In each phase, the groups in *Multilateral-Low* earn significantly more or not less than the groups in the other two institutions.

Table A5 compares contributions between high and low payoffs (for the same institution).

Table A5. Comparisons of contributions between *High* and *Low* payoff treatments

	Unilateral			*Majority*			*Multilateral*		
Phase	A	B	A & B	A	B	A & B	A	B	A & B
I	*High ≈ Low*	-	-	*High ≈ Low*	-	-	*High ≈ Low*	-	-
II	-	*High > Low***	*High ≈ Low*	*High ≈ Low*	*High ≈ Low*	*High ≈ Low*	*High ≈ Low*	*High ≈ Low*	*High ≈ Low*
III	*High ≈ Low*	*High > Low**	*High > Low**	*High ≈ Low*	*High > Low**	*High > Low**	*High ≈ Low*	*High ≈ Low*	*High ≈ Low*
IV	*High ≈ Low*	*High > Low***	*High > Low***	*High ≈ Low*	*High ≈ Low*	*High ≈ Low*	*High ≈ Low*	*High ≈ Low*	*High > Low***
V	*High ≈ Low*	*High > Low***	*High > Low**	*High ≈ Low*	*High ≈ Low*	*High > Low***	*High ≈ Low*	*High ≈ Low*	*High ≈ Low*

Notes: ≈ indicates no significant difference, < or > indicate statistically significant difference according to a two-sided MWW test. Significance levels: * P < 0.1, ** P < 0.05, *** P < 0.01.

Unilateral: Contributions never differ significantly when groups play game A. When groups play game B, under the single rule, in all phases, contributions are significantly higher with high payoffs than with low payoffs.

Majority: Contributions never differ significantly when groups play game A. When groups play game B, contributions are significantly higher with high payoffs than with low payoffs in phase III.

Multilateral: There are no significant differences in contributions between high payoffs and low payoffs, neither for game A nor for game B. Contributions over both games are significantly higher with high payoffs than with low payoffs in phase IV because more groups play game B with high payoffs.

We now present some helpful figures. The bars in Figure A1 show the voting behavior of individuals as a percent of total. The red lines show the percent of groups that play B and the green lines show the percent of groups that both play B and coordinate perfectly.

In Figures A2-A7 we show (on the left side) contributions and (on the right side) payoffs for every group in every treatment. (In our paper, we only show payoffs for the two Unilateral and two Multilateral treatments.)

Figure A1. Voting behavior and game choice (shares of individuals or groups)

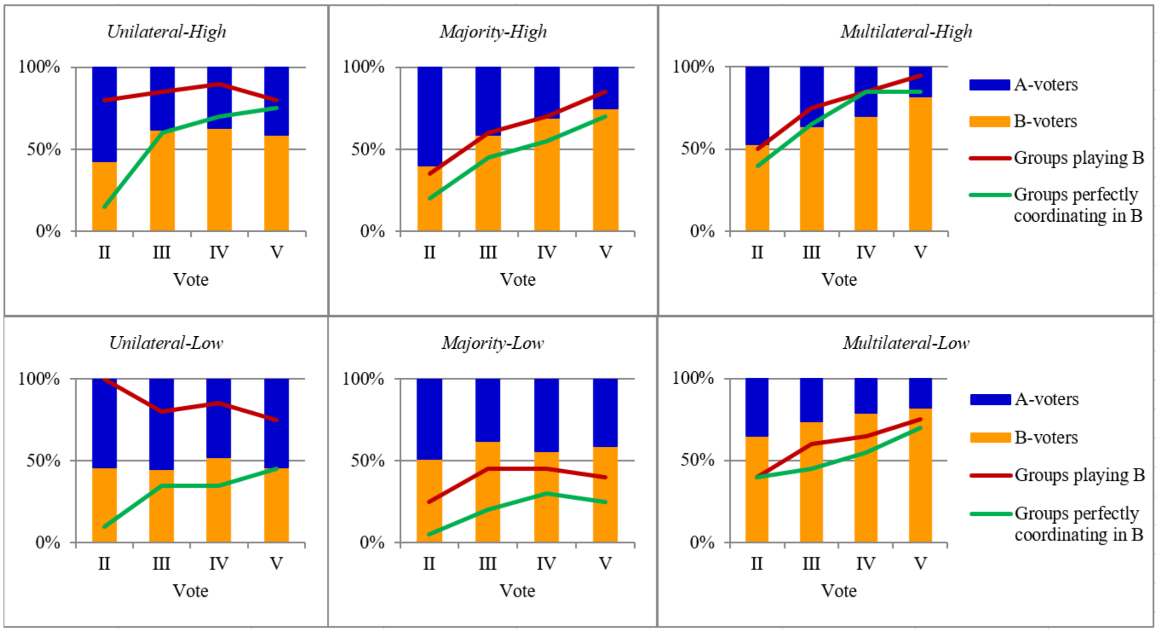


Figure A2. Contributions (left) and payoffs (right) by group for *Unilateral-High*

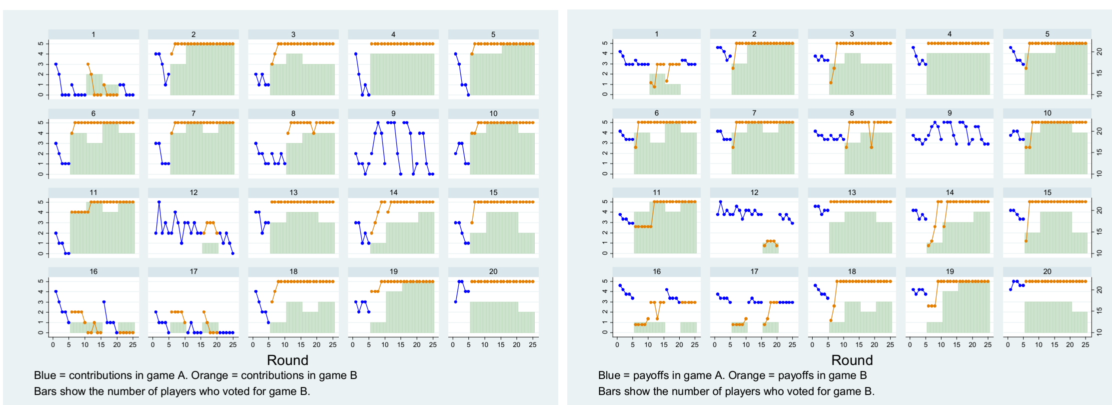


Figure A3. Contributions (left) and payoffs (right) by group for *Majority-High*

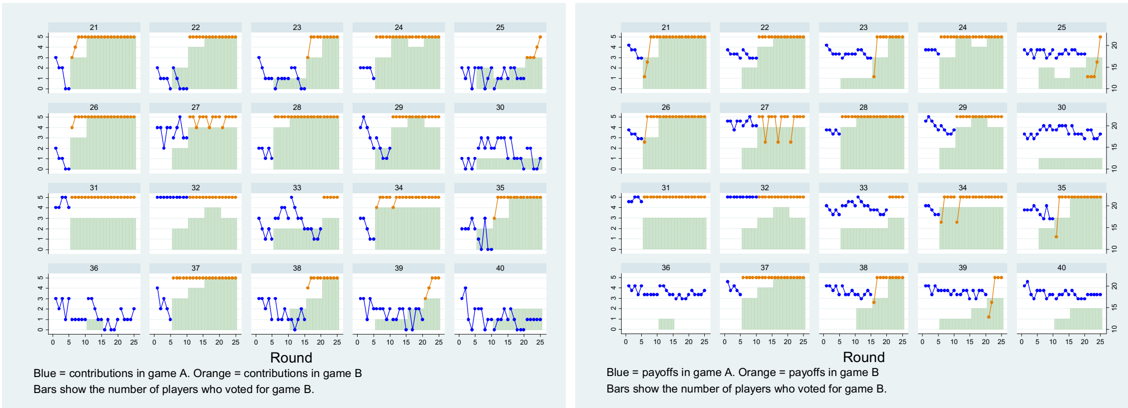


Figure A4. Contributions (left) and payoffs (right) by group for *Multilateral-High*

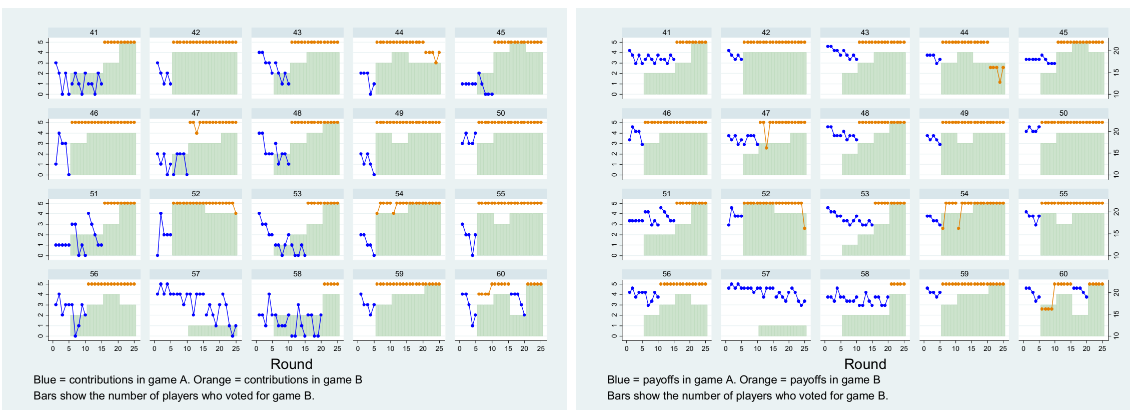


Figure A5. Contributions (left) and payoffs (right) by group for *Unilateral-Low*

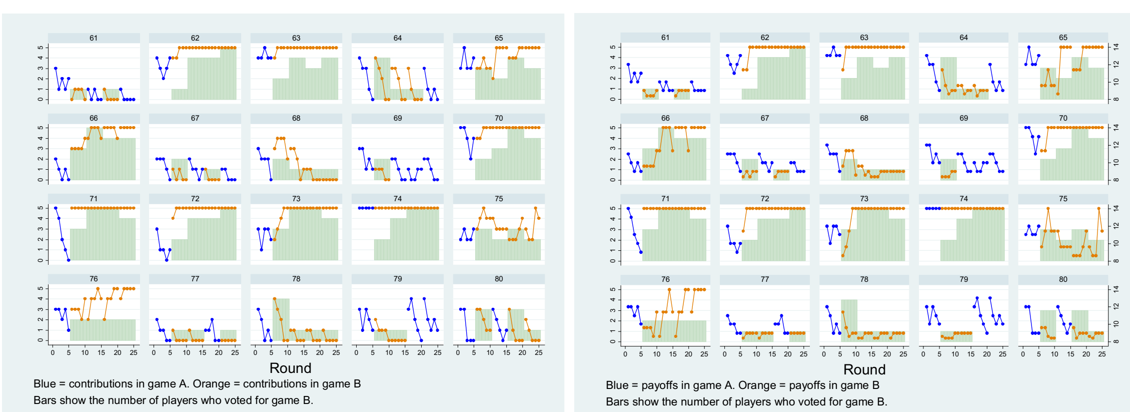


Figure A6. Contributions (left) and payoffs (right) by group for *Majority-Low*

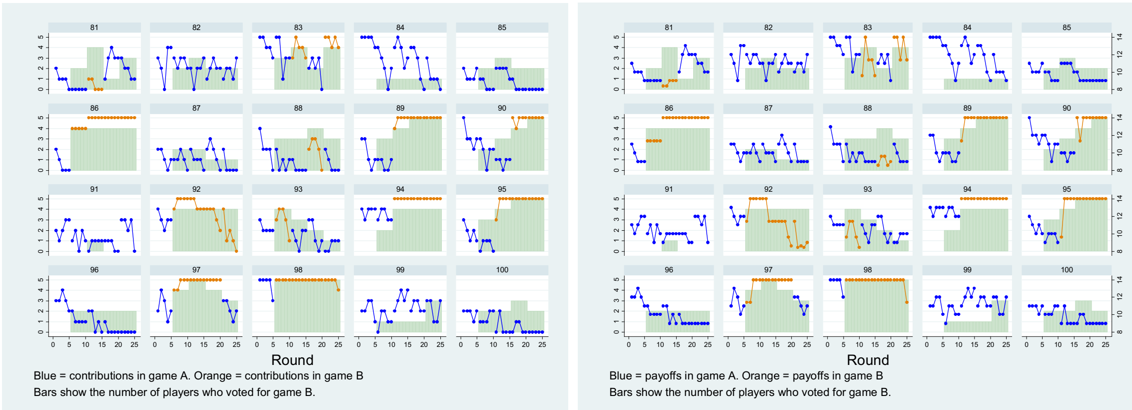


Figure A7. Contributions (left) and payoffs (right) by group for *Multilateral-Low*

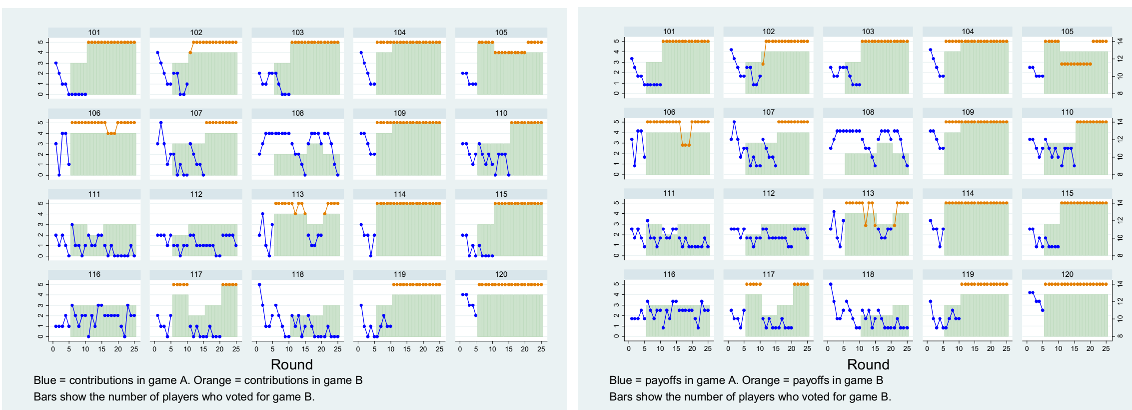


Table A6 provides further evidence of the role of beliefs in the form of probit regression results for individuals' voting decisions in the first voting phase (that is, in Phase II). As all groups were forced to play game A in Phase I, an individual's contribution decision in the first round of this phase provides a measure of his or her unconditional cooperativeness. Similarly, from the perspective of each individual, the contribution decisions of his/her co-players in this same round provide a measure of *their* cooperativeness—a measure that possibly influences each individual's beliefs about how his/her co-players will play in the remaining rounds. The regression results show that, in the *Unilateral* treatments, more cooperative individuals were more likely to vote for B than less cooperative individuals but others' cooperativeness did not significantly affect how each player voted. By contrast, in the *Multilateral* treatments, more cooperative individuals were not more likely to vote for B than less cooperative individuals, but others' cooperativeness did have a significant effect on how individual players voted. In *Multilateral*, individuals were more likely to vote for B if their co-players were relatively uncooperative. In short, the players who had reason to expect that cooperation would falter in A were more likely to vote for B.

Table A6. Regression results on the first voting decision by treatment

Variables	*Unilateral-High*	*Multilateral-High*	*Unilateral-Low*	*Multilateral-Low*
	First vote	First vote	First vote	First vote
Own cooperation first round	-0.191*	-0.124	-0.173*	-0.130
	(0.103)	(0.102)	(0.102)	(0.0972)
Others cooperation first round	-0.140	0.560***	0.146	0.334*
	(0.271)	(0.203)	(0.231)	(0.210)
Observations	100	100	100	100

Note: Numbers show marginal effects or discrete changes from probit regression models; standard errors in parentheses; level of significance: *** p<0.01, ** p<0.05, * p<0.1. Definition of variables: First vote = 1 if individual voted for A in phase 2, 0 otherwise; Own cooperation first round = 1 if individual cooperated in the first round of the first phase, 0 otherwise; Others cooperation first round = average cooperation of the co-players in the first round of the first phase.

Tables A7 and A8 show probit regression results on individuals' cooperation decision in the first round of the phase when game B or game A has been chosen by the group for the first time.

When B is chosen for the first time, the individuals' voting decision has a significant effect on the cooperation decision: subjects who have voted for game B are more likely to cooperate in game B. This effect is significant for *Unilateral* and *Majority* with both *High* and *Low* payoffs (of course, this happens by design when the institution is *Multilateral*). The number of A-voters in a group has the expected sign (negative: with fewer A-voters in the group an individual is more likely to cooperate) but the effect is only significant in *Unilateral-High*. Own cooperativeness has a significantly positive effect in *Unilateral-High* and *Unilateral-Low*. Others' cooperativeness has a positive effect in *Unilateral-Low*.

Table A7. Regression results on the first cooperation decision in *High* payoff treatments

	Game B chosen for the first time			Game A chosen for the first time		
	Unilateral-High	*Majority-High*	*Multilateral-High*[1]	*Unilateral-High*[1]	*Majority-High*	*Multilateral-High*
VARIABLES	Cooperate	Cooperate	Cooperate	Cooperate	Cooperate	Cooperate
Voted A	-0.290***	-0.181**			0.115	-0.115
	(0.106)	(0.0883)			(0.165)	(0.159)
Sum of A-votes	-0.0778*	-0.0747			-0.104	0.0726
	(0.0419)	(0.0599)			(0.0862)	(0.133)
Own cooperation first round	0.247**	0.0457	-0.00240	0.563***	0.563***	0.438***
	(0.103)	(0.0454)	(0.0284)	(0.197)	(0.197)	(0.138)
Others cooperation first round	0.188	0.205	-0.0323	0.163	0.163	0.303
	(0.269)	(0.130)	(0.0513)	(0.487)	(0.487)	(0.297)
Observations	95	85	95	30	65	55

Note: Numbers show marginal effects or discrete changes from probit regression models; standard errors in parentheses; level of significance: *** p<0.01, ** p<0.05, * p<0.1. Definition of variables: Cooperate = 1 if individual cooperated in the first round of the phase when game B or game A was chosen for the first time; Voted A = 1 if individual voted for game A in that phase, 0 otherwise; Sum of A-votes = number of group members who voted for game A in that phase; Own cooperation first round = 1 if individual cooperated in the first round of the first phase, 0 otherwise; Others cooperation first round = average cooperation of the co-players in the first round of the first phase. [1]The variables "Voted A" and "Sum of A-votes" are omitted if they predict success perfectly (game B in Agreement) or lack variation (game A in Single).

Table A8. Regression results on the first cooperation decision in *Low* payoff treatments

	Game B chosen for the first time			Game A chosen for the first time		
	Unilateral-Low	*Majority-Low*[2]	*Multilateral-Low*[1]	*Unilateral-Low*[2]	*Majority-Low*	*Multilteral-Low*
VARIABLES	Cooperate	Cooperate	Cooperate	Cooperate	Cooperate	Cooperate
Voted A	-0.355***	-0.389**			0.0624	-0.0513
	(0.117)	(0.159)			(0.123)	(0.126)
Sum of A-votes	-0.0926				-0.0383	0.111
	(0.0665)				(0.0763)	(0.0857)
Own cooperation first round	0.407***	0.0597		0.255	0.478***	-0.0442
	(0.118)	(0.138)		(0.196)	(0.123)	(0.124)
Others cooperation first round	0.902***	0.295		0.508	0.548**	-0.204
	(0.284)	(0.268)		(0.552)	(0.226)	(0.295)
Observations	100	55	4	35	85	70

Note: Numbers show marginal effects or discrete changes from probit regression models; standard errors in parentheses; level of significance: *** p<0.01, ** p<0.05, * p<0.1. Definition of variables: Cooperate = 1 if individual cooperated in the first round of the phase when game B or game A was chosen for the first time; Voted A = 1 if individual voted for game A in that phase, 0 otherwise; Sum of A-votes = number of group members who voted for game A in that phase; Own cooperation first round = 1 if individual cooperated in the first round of the first phase, 0 otherwise; Others cooperation first round = average cooperation of the co-players in the first round of the first phase. [1]For *Agreement-Low* and game B, regression is not possible due of lack of variation (almost all cooperated when game B was played the first time). [2]Some variables had to be omitted because they lack variation.

When game A has been chosen for the first time, the voting decision does not have a significant effect on the cooperation decision for *Majority* and *Multilateral* with both *High* and *Low* payoffs. In other words, A-voters and B-voters do not behave significantly different in game A. (Of course, this effect cannot be studied in *Unilateral* because all players in the group have to vote for A so that A is chosen.) Own cooperativeness has a positive effect on the cooperation decision in game A under all institutions with *High* payoffs. With *Low* payoffs, own cooperativeness has a significant positive effect only in *Majority-Low*.