Climate Leadership by Conditional Commitments*

Leif Helland† Jon Hovi‡ Håkon Sælen§

Abstract

In the 2015 Paris climate change agreement, each Party sets its own mitigation target by submitting a so-called Intended Nationally Determined Contribution (INDC). An important question is whether including conditional components in INDCs might enhance the agreement's effectiveness. We report the results of a closely controlled laboratory experiment that might help answer this question. The experiment investigates how two factors influence the effectiveness of leadership by (intrinsically) conditional commitments. We find that it may help if the leader's conditional promise is credible, that is, if followers have reason to believe that fulfilling the leader's condition will actually cause the leader's promise to be implemented. Moreover, it may also help if the leader, by implementing its conditional promise, is able to influence the followers' welfare substantially. Importantly, however, for both factors we find a significant effect only if the leader does not reap disproportionate gains from the group's collective efforts

Keywords: Laboratory experiment; conditional commitments; leadership; climate agreements

JEL Classification: C72; C92; F55; F64; H41

*Previous versions of this paper were presented at the International Studies Associationís Annual Convention in Toronto, 26 March 2014 (panel WC30), at the Oslo Academy of Global Governance workshop, 25 September 2014, and at the 9th NCBEE Conference in Aarhus, 26 September 2014. We are indebted to the participants in and attendants of these events and to Todd Cherry and Xinyuan Dai for helpful comments. The project was funded by the Department of Political Science at the University of Oslo, and grant 212996/F10 from the Norwegian Research Council.

†Department of Economics and Center for Experimental Studies and Research (CESAR), BI Norwegian Business School; leif.helland@bi.no (corresponding author)

‡Department of Political Science, University of Oslo.

§Center for International Climate and Enviromental Research Oslo (CICERO)

1 Introduction

In the 2015 Paris agreement under the UN Framework Convention on Climate Change, each Party sets its own mitigation target by submitting a so-called Intended Nationally Determined Contribution (INDC). The draft negotiation text includes competing suggestions regarding whether INDCs may contain conditional components (UNFCCC 2015, paragraph 26.1). If permitted, conditions might be expressed in two main ways. Intrinsic conditions would make own commitments conditional on other countries' mitigation efforts. In contrast, extrinsic conditions would make own commitments conditional on other countries' non-mitigation efforts, such as financial and technological support. An important question is whether intrinsic conditions, extrinsic conditions, or both, might enhance the effectiveness of the 2015 Paris agreement.

While extrinsic conditions have figured prominently in the UNFCCC in relation to action by developing countries,[1] intrinsic conditions – the focus of this paper – have less of a record in climate negotiations. However, as part of their 2020 pledges under the Cancun agreement, the European Union and Norway promised to cut emissions an additional 10% conditional on strong mitigation commitments by other Parties (UNFCCC 2011b). These intrinsically conditional commitments apparently had little (if any) effect on other countries; hence, they were not implemented. The EU countries and Norway have not, at the time of writing, presented conditional commitments for the 2015 agreement, but they have left the door open for doing so later in the process by formulating their INDCs as at least 40% reductions.

Moreover, intrinsically conditional commitments constitute a central element in David Victor's (2011) club approach to climate cooperation outside the UNFCCC. Victor (2011) suggests that cooperation should begin with agreements between small groups of enthusiastic countries. The backbone of his approach is a series of conditional offers, whereby enthusiastic countries would outline what they are willing and able to do, conditional on what other enthusiastic governments offer and implement. Moreover, reluctant countries would be enticed to join via "exclusive and contingent" measures, such as preferential market access for club members.[2]

Hence, an important question concerning climate cooperation both inside and outside the UNFCCC process is whether and, if so, under which conditions countries might enhance cooperation through leadership based on intrinsically conditional commitments.

We present results from a closely controlled laboratory experiment that might help inform the conditionality debate. The basic preference configuration underlying the INDC process resembles the one found in one of the most widely studied games in experiments, the voluntary contribution mechanism game, also known as the public good game. We study a novel variation of this game to assess how two factors influence the effectiveness of leadership by intrinsically conditional commitments.[3] First, it may help if the leader's conditional promise is credible, that is, if followers have reason to believe that fulfilling the leader's stated condition will actually cause the leader's promise to be implemented. Second, it may also help if the leader has the ability to influence the followers' welfare substantially by implementing its conditional promise.

For both factors we find a significant effect only if the leader does not reap disproportionate gains from the group's collective efforts. This finding supports recent results from research on simultaneous-move public good games with punishment opportunities. These results show that

[1] For example, the Cancun agreement states that developing countries' mitigation shall be «supported and enabled by technology, financing, and capacity-building» (UNFCCC 2011a).

[2] Victor 2011 also suggests that agreements should (a) be nonbinding, (b) entail high flexibility concerning choice of policy strategies, and (c) focus on policies that governments actually control, rather than on emission levels (which in large part depend on factors beyond governmental control).

[3] These conditions resemble the corresponding conditions necessary for a threat to be effective (e.g., Schelling 1960). Weischer et al. (2011) elaborate on the conditions for a promise to be effective in the context of climate change.

efficient contribution norms do not easily evolve in groups where some members benefit significantly more than others from collective efforts. We also find that the outcome remains severely suboptimal even under favorable conditions. Thus, it seems that if leadership by conditional commitments is to bring about efficient mitigation levels, it must be supplemented by other measures.

To the best of our knowledge, our experiment is the first one to study leadership by conditional commitments systematically. The environment we study is highly stylized: The game's structure is public knowledge; the sequence of moves, the time horizon, time periods, payoffs, and contributions are all unambiguously defined; subjects can observe behavior without delay or noise; and all decisions are anonymous. This stylized environment only slightly resembles real-world settings where conditional commitments are or can be used. Thus, the external validity of our results should be checked through field studies—if and when relevant field data become available.

On the other hand, an experimental design built on a stylized environment has its advantages. Empirical field data from international negotiations are not only scant; they also suffer from familiar problems such as endogeneity; selection on unobservables; omitted third variables; and reverse causality. Experiments allow randomization over treatments and truly exogenous variation in the explanatory variables; hence, the conditions (if any) under which conditional commitments might be effective can be investigated by systematically manipulating the structure of interactions.

In section 2, we review relevant literature. In section 3, we provide some basic theory. In section 4, we outline our underlying model and the treatments. In section 5, we provide details about the experiment's design and implementation. In section 6, we present our results. Finally, in section 7, we conclude.

2 Related Literature

We review four strands of relevant literature. The first consists of theoretic (mostly game-theoretic) work on leadership in the form of unilateral emissions reductions. This strand offers little (if any) support for the conjecture that unilateral action will induce other countries to follow suit. Using a two-country model, Hoel (1991) demonstrates that if one country (the leader) undertakes unilateral emissions reductions, the other country (the follower, which is assumed to be motivated by self-interest) may well increase its own emissions.

The reason is that the leader's unilateral action diminishes the follower's marginal benefit of emissions reductions. Hoel also finds that unilateral emissions reductions may cause international climate change negotiations to result in an agreement with higher total emissions than if both countries act selfishly (in which case no unilateral action will occur).

More recent studies support Hoel's results. For example, Buchholz and coauthors (1998) find that other countries' free riding will likely offset unilateral efforts by one or a few countries. Thus, in their model (which closely resembles Hoel's) a coalition acting unilaterally can generate net benefits to its members only if it includes all major emitters.

Similarly, using a coalition model, Holtsmark (2013) shows that if one country were to announce ambitious and unconditional emissions reductions before international negotiations take place, the result would likely be higher global emissions than if the country were to make its pledge conditional on other countries' undertaking similar emissions reductions in return (Holtsmark 2013).[4]

[4] Common to these game-theoretic studies is that they ignore the possibility of "no-regret" options for reducing emissions of greenhouse gases. Ott and Oberthür (1999) suggest that a leader might cause global emissions reduc-

Lastly, using an incomplete-information model, Konrad and Thum (2014) find that a unilateral and unconditional commitment to reducing emissions diminishes the gains from global cooperation and hence makes it more difficult to reach an effective international agreement.

A second strand consists of game-theoretic work focusing on the prospects for transforming the climate change mitigation game from a social-dilemma game to a coordination game. The underlying assumption is that countries are much better at solving coordination games than they are at solving social-dilemma games. For example, Barrett (2003) shows how trade restrictions and technology standards might serve this function. Moreover, Barrett and Dannenberg (2012) find that a looming climate disaster with a known emissions threshold could create a coordination game. We show that a leader in a social-dilemma game might be able to create a coordination game for the followers through a properly designed conditional commitment.

The third strand consists of political science work that is more empirically oriented than the contributions in the first and second strands. While political scientists studying climate leadership and minilateralism have been more concerned with fairness than with effectiveness,[5] Skodvin and Andresen (2006) present a case study of the EU's attempt to exert leadership by saving the Kyoto Protocol after the US repudiation in 2001. They conclude that, although EU leadership was instrumental for Kyoto's entry into force, the resulting agreement was a mini-regime with "miniscule impact on climate change abatement". Based on a quantitative study, McLean and Stone (2012) arrive at similar conclusions: Having ratified the Kyoto protocol jointly, the EU countries ensured the protocol's entry into force by offering selective incentives (such as EU membership or support for WTO membership) to other countries.

Combining simulations with case studies, Underdal et al. (2012) focus specifically on leadership by conditional commitments. They argue that such leadership can work–but only under rather strict conditions. In particular, they find that successful leadership requires that two conditions be fulfilled: First, the leader must promise to undertake substantial additional emissions reductions if other countries fulfill the stated conditions. Second, the leader's promise must be credible, so that followers expect the leader to implement its promise of additional emissions reductions if (and only if) other countries fulfill the leader's stated conditions.

Finally, the fourth strand consists of a small but growing body of experimental research on leadership in public good games. It is well known that subjects' behavior in public-good experiments tends to deviate systematically from standard game-theoretic predictions, which are based on the assumptions of purely self-interested motivation and common knowledge of rationality. In particular, subjects in public-good experiments contribute and (when given the opportunity) punish substantially more than suggested by the stark zero-contribution and zero-punishment predictions of standard game theory. The reasons for these deviations have been extensively explored in the literature (see e.g., Chauduri 2011; Fehr and Gächter 2000; Kosfeld et al. 2009; McEvoy 2010; McEvoy et al. 2011; Ostrom 2000; and Ostrom et al. 1992).

For small groups (4 to 10 subjects) and sizable marginal per capita return on contributions (MPCR between 0.30 and 0.75) group size does not significantly affect contribution behavior. In contrast, the MPCR, controlled for group size, significantly influences contributions in small-group, high-MPCR settings (Isaach and Walker 1988).

In a recent study by Weimann and coauthors (2012), it is found that this relationship holds even for sizable groups (40 to 60 subjects) and for very low MPCRs (0.02 and 0.04). This finding indicates that small-group behavior in the lab is also relevant for large-scale problems where the marginal benefits of individual contributions to a public good are negligible, as is typically the case for global emissions reductions.

tions by demonstrating such options' attractiveness to other countries. However, this alleged effect would seem to presuppose that the leader has superior knowledge concerning no-regret options —a rather strong assumption.

[5] See e.g., Eckersley (2012); and Maltais (2014).

Cherry et al. (2005) and Fischer et al. (1995) find that heterogeneously endowed subjects contribute significantly less than homogenously endowed subjects do. In contrast, Reuben and Riedl (2013) find that both heterogeneous endowments and heterogeneous returns produce approximately a doubling of contributions relative to the contributions in homogeneous groups when no punishment is available. However, when punishments are introduced, the increase in contributions is substantially weaker with heterogeneous MPCRs than with heterogeneous endowments. The authors conclude that subjects converge on efficient contribution norms even when endowments differ, but not when subjects benefit unevenly from public-goods provision. According to Reuben and Riedl (2013), uneven benefits give rise to conflicting contribution norms.[6] Such conflicting norms hamper cooperation.

While most public-good experiments implement simultaneous moves, only a handful let one group member (the leader) make its contribution decision before the other group members (followers) do so. Güth and coauthors (2007) find that experiments with (unconditional) leadership trigger higher average contributions than standard public-good experiments with simultaneous moves do. This difference in contributions is statistically significant, yet substantially moderate. These results thus suggest that while unconditional leadership enhances cooperation, it comes nowhere near fully solving the underlying collective-action problem.[7]

Their conclusion is supported by Levati and coauthors (2007), who find that the effect of unconditional leadership is even weaker (but still significant) when the subjects' endowments differ and this difference is public knowledge.

Gächter and coauthors (2010) find that reciprocator types contribute significantly more than self-interested types in the role of leader.[8] A substantial part of this effect, however, is due to false consensus. Reciprocator types initially tend to overestimate the number of other reciprocators in the population and hence contribute substantially in the first round. However, they are disappointed when other followers' contributions are lower than expected. Disappointment due to false consensus may, at least partly, explain why average contributions are falling over time in the experiments such as the ones conducted by Güth and coauthors and Levati and coauthors.

In contrast, Rivas and Sutter (2011) find a substantial effect of leadership on contributions when leaders are permitted to self-select, rather than being allocated, into the leader role.[9] Moreover, with voluntary leadership, average contributions do not appear to be falling over time. These findings lend some support to the conjecture that enthusiastic leaders may make a difference. However, in the set-up of Rivas and Sutter leaders are not permitted to condition their contributions on follower behavior—which is the focus of our experiment.

Of the contributions reviewed here, only the contributions by Underdal and coauthors and by Holtsmark consider leadership by conditional commitments (as we do). Using Güth et al.'s unconditional leadership treatment as baseline, we introduce several novel treatments. These treatments aim at pinpointing the conditions under which leadership by conditional commitments can or will be effective. Our treatments introduce changes step by step, so that only a single experimental design element differs from one treatment to the next.

[6] Fisher et al (1995) find that a player's MPCR has a strong positive effect on that player's contribution, but find no effect of MPCR heterogeneity on group contributions.

[7] Compared to the baseline in which all subjects choose simultaneously, average contributions (over all periods and all groups) increase by 13.5 percentage points (from 40 percent in the baseline). In an additional treatment the leader is granted the right to exclude one member of the group from consuming the public benefits in the next period. This treatment increases average contributions by 39 percentage points compared to the baseline. Given our motivation, however, this sanctioning mechanism is not very interesting. A viable global climate is a pure public good, and excluding states from its benefits is not feasible.

[8] In this study the distribution of types is extracted using the strategy method (proposed by Selten 1967) prior to actual decision-making in the experiment.

[9] Compared to the simultaneous-choice baseline (with average contributions of 40 percent) voluntary leadership increases average contributions by almost 23 percentage points.

3 Model and Treatments

Consider a three-stage one-shot game where one player is randomly selected as leader (L), while the other $n-1$ players are followers (F). Each player is endowed with z_k units of a numéraire good (with $k = \{L, F\}$).

In stage one, the leader decides how much of its endowment to contribute to a public account for the group. In our eight main treatments (T3 through T10), the leader can also promise to top up its contribution in stage three, provided that the followers' average contribution exceeds a minimum specified by the leader.[10]

In stage two, followers—having observed the leader's contribution and conditional promise (if any)—decide simultaneously how much of their endowment they will contribute to the public account; thus, player i's contribution c_i must satisfy $c_i \in [0, z_k]$. Once made, the followers' aggregate contributions are observed by the n players.[11]

In stage three, the leader's contribution can be increased (unless the leader contributed its entire endowment in stage one). In some treatments, the leader is free to choose whether it will top up its contribution and, if so, by how much. In other treatments, a computer program automatically implements the leader's conditional promise whenever followers fulfill the leader's condition. All treatments except T1 include this third stage (T1 consists of stages one and two only, and replicates Güth et al. 2007).

Contributions are multiplied by a factor greater than unity and less than the number of players in the group—either evenly or relative to the players' endowments (see Table 1). Unless a contribution can be pivotal for increasing one or more other players' contributions, it is a strictly dominant strategy to contribute zero units to the public account (assuming rationality, self-interest, and complete information).

Our 10 treatments were designed to study under what conditions leadership by conditional commitment will effectively enhance followers' contributions (Table 1). We are particularly interested in the effects of (1) giving the leader the possibility to explicitly state its conditions for topping up; (2) making the leader's conditional promise binding (i.e., fully credible); (3) expanding the leader's endowment; and (4) increasing the leader's MPCR.[12]

Player i's payoff π_i equals:

$$\pi_i = z_k - c_i + \alpha_k \sum_{i=1}^{n} c_i$$

Here the first right-side term (z_k) represents player i's endowment, the second (c_i) represents player i's contribution,[13] and the third represents player i's benefit from its own and others' contributions, with α_k representing the MPCR of each unit contributed to the public good. In all treatments, $n = 4$ and $c_F = 100$. The values of z_L, α_L, and α_F vary across treatments (see

[10] In T1 and T2, the leader cannot make such a conditional promise.

[11] The sequence of moves in actual climate negotiations is not fixed, but endogenous. If followers move in a pre-determined sequence and promises are non-binding, a unique equilibrium exists in which no player contributes (by backwards induction). If promises are binding it exists a cooperative equilibrium in addition to the non-cooperative one. With a pre-determined sequence of moves and binding promises, however, there is a unique way to play the cooperative equilibrium. This is in contrast to the case were followers move simultaneously, and were there may be multiple ways to play the cooperative equilibrium (see below).

[12] Interpreting (3) and (4) in a climate context: A leader can be "big" in two ways; by having a large endowment, which can be interpreted as having a large capacity to emitt; and having a large marginal benefit of a given abatment, which can be interpreted as having a large population benefiting from it.

[13] For the leader, c_i represents the sum of its contribution in stage 1 and its contribution in stage 3.

Table 1). We make sure to keep the social return on contributions to the public good constant as we vary α_L and α_F over treatments.[14]

Treatment	Short description	Detailed description	z_L	α_L	α_F
T1	Baseline/control	Standard sequential-simultaneous public-good game. Leader moves first and followers move simultaneously after having observed the leader's contribution.	100	0.4	0.4
T2	Implicit conditionality	As T1, except that leader can top up (i.e., make a second contribution decision) in stage three, after having observed the followers contributions.	100	0.4	0.4
T3	Explicit conditionality, nonbinding promise	As T2, except that leader can make a nonbinding, conditional promise to top up. The condition is that the followers' average contribution must exceed a minimum chosen by the leader.	100	0.4	0.4
T4	Explicit conditionality, binding promise	As T3, except that the leader's promise is binding, in the sense that if the leader's stated condition is fulfilled, then the promise is automatically implemented by the computer.	100	0.4	0.4
T5	Explicit conditionality, binding promise, public gains shared unevenly	As T4, except that the leader's share of the gains from the public account equals twice that of a follower.	100	0.64	0.32
T6	Explicit conditionality, binding promise, big leader, public gains shared evenly	As T4 except that the leader's endowment equals twice that of a follower.	200	0.4	0.4
T7	Explicit conditionality, binding promise, big leader, public gains shared unevenly	As T4, except that the leader's endowment equals twice that of a follower and that the leader's share of the gains equals twice that of a follower.	200	0.64	0.32
T8	Explicit conditionality, nonbinding promise, public gains shared unevenly	As T5, except that the leader's promise is nonbinding.	100	0.64	0.32
T9	Explicit conditionality, nonbinding promise, big leader, public gains shared evenly	As T6, except that the leader's promise is nonbinding.	200	0.4	0.4
T10	Explicit conditionality, nonbinding promise, big leader, public gains shared unevenly	As T7, except that the leader's promise is nonbinding.	200	0.64	0.32

Table 1: *Treatments*

[14] Specifically $\alpha_L + 3\alpha_F = 1.6$ both when $\alpha_L = \alpha_F = 0.4$ and when $\alpha_L = 0.64$ and $\alpha_F = 0.32$.

We begin by considering a situation in which the assumptions of what we might call the "standard model" apply: In this situation, it is common knowledge that all n players are rational and purely self-regarding. Based on these assumptions, what will be the game's subgame-perfect equilibrium? The answer depends on whether the leader's promise is binding.

First, consider the case where the leader's promise is nonbinding, so that in stage three, the leader is free to choose whether it will keep or violate its promise (if any) from stage one. Using backward induction we find that in stage three, the leader will contribute zero. The reason is that the marginal cost of contributing one unit is 1, whereas the marginal private benefit of contributing one unit is only α_L (< 1). Moreover, contributing a positive amount in stage three cannot influence followers' contributions, simply because followers have no decisions to make after stage two (in the one-shot game).

In stage two, no follower will make a contribution, because $\alpha_F < 1$ and because followers anticipate that, regardless of their decisions, the leader will contribute zero in stage three.

Finally, in stage one the leader will contribute nothing, because $\alpha_L < 1$ and because the leader anticipates that, regardless of the leader's stage-one contribution (and promise, if any), no follower will make a contribution in stage two.

It follows that for T1 and T2 (in which the leader can make no promise at all) as well as for T3, T8, T9, and T10 (in which the leader can make only a nonbinding promise), the unique subgame-perfect equilibrium of the standard model is that all n players contribute nothing.[15] Thus, in these treatments each player's equilibrium payoff equals its endowment z_k. Because $\alpha_k > 1/n$ in our design, this subgame-perfect equilibrium is Pareto dominated by the non-equilibrium outcomes wherein all players contribute their entire endowment. Note that many such outcomes exist, because the leader can divide its contribution of z units between stage one and stage three in many different ways.[16]

Backward induction shows that in a finitely repeated game, the stage-game equilibrium will be played in every period. Thus, in T1, T2, T3, T8, T9, and T10 the subgame-perfect equilibrium in the finitely repeated game is that all players contribute zero units in every period.

Next, turn to the case where the leader's promise is binding (T4 through T7). For this case, our experimental design makes it public knowledge that if followers fulfill the leader's condition, then the leader's promise to make an additional contribution in stage three would be automatically implemented by the computer. In all treatments where the leader's promise is binding, cooperative equilibria exist. In particular, the leader can—by choosing its stage-one contribution and promise appropriately—create a coordination game for the followers. Denote the leader's conditional contribution b and its minimum requirement for the followers' average contribution c^*. Follower i's return from contributing to the public account will then equal:

$$c_i(\alpha_F - 1) \; if \; c_i < (n-1)c^* - \sum_{j \neq i} c_j$$

and

$$c_i(\alpha_F - 1) + b\alpha_F \; if \; c_i \geq (n-1)c^* - \sum_{j \neq i} c_j$$

[15] What would the leader promise in equilibrium? Because the promise is nonbinding, it is costless to make any promise as well as to violate it (cheap talk). Any promise is thus consistent with equilibrium behavior.

[16] When the leader must choose an integer between 0 and 100 (as in our experiment), exactly 101 such Pareto-optimal outcomes exist.

where c_j is follower j's contribution. Notice that follower i's return function shifts vertically at the point where i's contribution is pivotal for triggering implementation of the leader's conditional contribution. Denote this point c_i^*. The return function is non-negative if $c_i^*(\alpha_F - 1) + b\alpha_F \geq 0$. Solving for b gives:

$$b \geq \frac{(1-\alpha_F)}{\alpha_F} c_i^* \tag{1}$$

When condition (1) holds, a follower has no incentive to deviate unilaterally from c_i^*; hence, c_i^* constitutes a best reply, given the other players' contributions (and the leader's promise). Dropping the subscript on c^*, condition (1) gives the combinations of b and c^* with which the leader will create a coordination game for the followers. All combinations of follower contributions that exactly meet the leader's stated minimum constitute equilibria of the stage 2 game; however, zero contributions by all players is also an equilibrium. The positive-contribution equilibria Pareto dominate the zero-contribution equilibrium whenever condition (1) is a strict inequality. However, zero contribution is the maximin strategy for followers.

Given that followers coordinate to meet c^*, the leader's marginal benefit with regard to c^* is $\alpha_L(n-1)$ while the marginal cost is $1 - \alpha_L$, which implies positive net marginal benefit as long as

$$\frac{\alpha_L}{(1-\alpha_L)} > \frac{1}{(n-1)} \tag{2}$$

This condition is always satisfied in our experiment. Furthermore, when followers fail to coordinate, the leader's marginal benefit and marginal cost with regard to c^* are both zero. The leader will maximize its payoff by maximizing c^* subject to its own budget constraint, the followers' budget constraint, and condition 1. Solving yields

$$c^* = \min\left[z_L \times \frac{\alpha_F}{(1-\alpha_F)}, z_F\right] \tag{3}$$

$$b = \frac{(1-\alpha_F)}{\alpha_F} c^* \tag{4}$$

The numeric equilibrium solutions in terms of the leader's contribution and the required average followers' contribution are listed in Table 2 for the treatments with binding promises.

Treatment	Equilibria $(c_L, \bar{c}_F)$	Percent of potential
T4	$(0,0)$ and $(100,66)$	74.5
T5	$(0,0)$ and $(100,47)$	60.3
T6	$(0,0)$ and $(150,100)$	90.0
T7	$(0,0)$ and $(200,94)$	96.4

Table 2: *Equilibrium leader contribution and equilibrium average follower contribution in treatments with binding promise*

Note that more than two equilibria typically exist: unless the leader's condition requires all followers to contribute their entire endowment, the followers can take on the costs involved in satisfying the leader's stated condition in (many) different ways. Given this coordination problem, it is by no means obvious that the followers will manage to settle on any particular positive-contribution equilibrium. Treatment 6 provides an exception to the multiplicity of positive contribution equilibria. In this treatment followers are required to contribute their entire endowment in equilibrium, reducing the number of positive contribution equilibria to one. To ensure that the positive-contributions equilibrium is Pareto dominant, the leader must set the ratio of b to c^* higher than implied by equations (3) and (4).

The last column in Table 2 provides the sum of contributions in the positive contributions equilibrium, as a percentage of the sum of endowments. This percentage can be taken as a measure of efficiency, since the payoffs are increasing linearly in the sum of contributions. As can be seen, having high levels of efficiency in public goods provision as an equilibrium phenomenon requires high leader endowments and access to a commitment technology.

4 Design and Implementation

As explained in the previous section, we ran 10 treatments (including the control treatment). Each treatment consisted of 16 periods. To avoid "envy effects," we let each subject act as leader for four (subsequent) periods. Which subject acted as leader in which four periods was determined randomly.[17]

We recruited a total of 408 subjects for the experiment, 176 subjects from the general student population at BI Norwegian Business School and 232 subjects from the general student population at Appalachian State University, Boone. The number of groups included in a session varied from 3 to 7. No subject participated in more than one session. We ran a total of 21 sessions for the experiment, striving to balance the US and Norwegian sessions over the 10 treatments. The sessions were conducted between May 2013 and May 2014.

We implemented a partner design in which the four-subject groups were formed randomly at the beginning of each treatment and remained constant for that treatment's 16 periods. Subjects only received feedback about behavior in own group. In each period, all subjects received information about the leader's contribution prior to entering stage 2. After followers had made their (simultaneous) contribution decisions in stage 2, all subjects were informed about these follower contributions. When relevant (T2-T10), all subjects were informed about the leaders stage three contribution after it had been made. Finally in the end of each period all subjects got a feedback in the form of a statistics covering decisions and payoffs in the current and all previous periods. Subjects' anonymity was preserved throughout.

All sessions were computerized, and the experiment was programmed in z-tree (Fischbacher 2007). In each session the administrator, having seated the subjects at randomly drawn cubicles in the lab, distributed the instructions and read them aloud. The session began after subjects had answered a set of control questions designed to ensure they understood the payoff structure. Each session lasted about one hour. In the experiment an Experimental Currency Unit (ECU) was used. The instructions made the exchange rate from ECU to USD or NOK public knowledge.

Subjects received their earnings, which averaged around USD 40 / NOK 250, in cash and privately, at the end of the session, which lasted on average around one hour.

[17] Literature on the possible effects of role reversal is scarce; however, the scant evidence that does exist seems to point in the direction of no significant effect. For example, Hall (2013) finds no significant difference in behavior between role-reversal and single-role protocols for a trust game.

5 Results

We present our results by way of (1) bar charts with corresponding non-parametric tests of differences and (2) regressions. We are particularly interested in how the average follower contribution varies by treatment. In addition, we study how often the leader creates a coordination game for the followers and whether followers are able to coordinate by meeting (or exceeding) the leader's stated condition.

Average follower contribution: non-parametric tests

Figure 1 shows the average follower contribution and the average leader contribution for all of our 10 treatments. Six main features stand out.

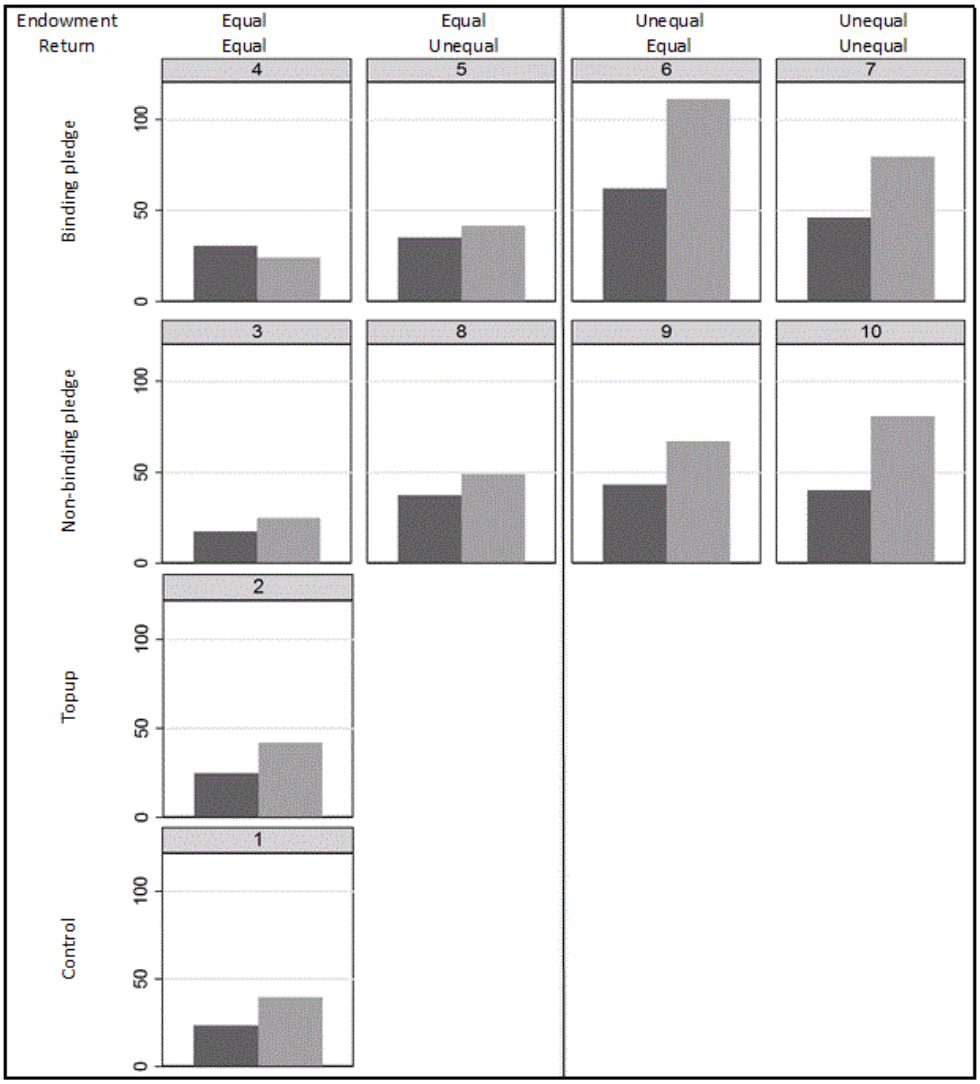

Figure 1: *Average leader contribution (light grey) and average follower contribution (dark grey). All treatments and all sessions.*

First, the average follower contribution varies considerably across treatments. In particular, it is more than three times higher in treatment 6 (the maximum) than in treatment 3 (the minimum). Thus, the variables defining our treatments seem to influence the followers' behavior.

Second, the average leader contribution also varies considerably across treatments. Thus, the variables defining our treatments seem to influence the leader's behavior as well.

Third, the average follower contribution is positively correlated with the average leader contribution. This finding suggests that the leader's behavior influences the followers' behavior (and possibly vice versa).

Fourth, the average follower contribution is not higher in treatments 2 and 3 than in treatment 1. Thus, giving the leader the opportunity to top up or to top up and make a conditional promise does not by itself enhance public-goods provision.

Fifth, and consistent with the equilibrium of our model, the average follower contribution is higher in treatments where (1) the leader has a large endowment and (2) implementation of the leader's promise is automatic, than in treatments with neither of these two features. For example, compare treatment 6 (z_L=200 and automatic implementation) with treatment 3 (z_L=100 and voluntary implementation). The impact of coordination is analyzed further below.

Finally, only one treatment (treatment 6) displays an average follower contribution higher than 50% of the endowment. Thus, public-goods provision remains moderate even under quite favorable conditions.

The main result is summarized in result 1:

Result 1 *Leadership by conditional commitment enhances public-goods provision under some conditions, yet falls substantially short of solving the collective-action problem faced by the subjects in our experiment.*

We now proceed by formally testing differences between treatments with explicit conditionality (T3–T10). In particular, we focus on followers' average contributions conditioned on: (1) whether the leader's promise is binding (automatic implementation) or not binding (implementation left to the leader's discretion); (2) whether the leader's endowment is large ($z_L = 200$) or small ($z_L = 100$); and (3) whether the benefits from the public account are shared evenly ($\alpha_L = 0.4$) or unevenly ($\alpha_L = 0.64$).

Binding promise:				
Yes	No	Endowment	Returns	p-value
T6 (8)	T9 (10)	Uneven	Even	.110
T4 (12)	T3 (10)	Even	Even	.056
T5 (8)	T8 (8)	Uneven	Uneven	.916
T7 (12)	T10 (10)	Even	Uneven	.391

Table 3.1: *WRS tests of differences in the average follower contribution by credibility of promise (number of groups)*

Endowment:				
Uneven	Even	Binding:	Returns	p-value
T6 (8)	T4 (12)	Y	Even	.002
T9 (10)	T3 (10)	N	Even	.019
T7 (12)	T5 (8)	Y	Uneven	.143
T10 (10)	T8 (8)	N	Uneven	.790

Table 3.2: *WRS tests of differences in average follower contribution by endowment (number of groups)*

Returns:				
Even	Uneven	Endowment:	Binding	p-value
T6 (8)	T7 (12)	Uneven	Y	.031
T4 (12)	T5 (8)	Even	Y	.440
T9 (10)	T10 (10)	Uneven	N	.821
T3 (10)	T8 (8)	Even	N	.002

Table 3.3: *WRS tests of differences in average follower contribution by returns (number of groups)*

First, by comparing treatments vertically in Figure 1 (T4 vs. T3; T5 vs. T8; T6 vs. T9; T7 vs. T10), we appreciate that the effect of binding promises is positive except when benefits are unevenly distributed. Thus, being able to set forth binding promises seems to help leaders induce followers to contribute, but only if the proceeds from the public account are distributed evenly. Wilcoxon rank-sum (WRS) tests confirm this finding.[18] The results in Table 3.1 show that one effect (T4 vs. T3) is close to significance at the 5% level, while another (T6 vs. T9) is close to significance at the 10% level. Thus, binding promisess seems to have an effect on average follower contributions only when benefits are evenly distributed.

Second, by comparing treatments that vary only concerning leader endowments in Figure 1 (T3 vs. T9; T8 vs. T10; T4 vs. T6; and T5 vs. T7), we find that the size of the leader's endowment has a positive effect on the average follower contribution in all four cases. Table 3.2, however, reveals that the effect is marginal at best when benefits are unevenly distributed. WRS tests confirm that big endowments help leaders induce followers to contribute only if proceeds are shared evenly.

Finally, by comparing treatments that vary only concerning how benefits are distributed in Figure 1 (T4 vs. T5; T6 vs. T7; T3 vs. T8; T9 vs. T10), we find that increasing the leader's MPCR (while simultaneously decreasing the MPCRs of followers) appears to have a positive effect on the average follower contribution in two of the four cases (4 vs. 5; 3 vs. 8), but a negative effect in the other two cases (7 vs. 6; 9 vs. 10). Table 3.3. reveals that one effect in each direction is significant at the 5% level. Thus, based on Figure 1 and the WSR tests the isolated effect of unevely distributed benefits seems indeterminate.

These three findings are summarized in Result 2:

Result 2 *Endowment and binding promises interact with returns concerning their effect on followers' average contribution.*

We now proceed to study our treatment effects using regressions. Her we pay particular attention to interaction effects.

Average follower contribution: regressions

In the regressions we use individual decisions as the unit of analysis with individual random effects, and clustering of standard errors at the group level to control for within-group interactions.[19]

[18] We conducted pairwise Wilcoxon rank-sum tests of differences over treatments, using group averages over the 16 periods of play as observations. These tests take account of group differences caused by variations in strategic interaction and learning over time.

[19] We performed a series of robustness tests using alternative model specifications. Using group averages as the dependent variable (rather than individual decisions), with group random effects and standard errors clustered

Table 4 reports the results of two GLS regressions. Model 1 includes our five institutional variables: Top-up (scores 1 if leader can top up, 0 otherwise), Promise (scores 1 if leader can make a promise, 0 otherwise), Endowment (scores 1 if $z_L = 200$, scores 0 otherwise), Binding (scores 1 if the promise is implemented automatically, 0 otherwise), and Returns (scores 1 if $\alpha_L = 0.64$, 0 otherwise). It also includes the control variable Lab (scores 1 for US sessions, 0 for sessions at author 1's institution).

	Model 1			Model 2		
	Coefficient	t-value	p-value	Coefficient	t-value	p-value
Binding	7.66	2.09	0.036	8.07	2.48	0.013
Endowment	16.91	4.60	0.000	10.76	3.07	0.002
Returns	1.83	0.49	0.627	0.59	0.17	0.862
Top up	4.26	0.74	0.462	3.82	0.79	0.430
Promise	-0.71	-0.12	0.903	1.77	0.35	0.724
Lab	12.34	3.87	0.000	10.45	3.68	0.000
Unconditional c_L				0.12	4.17	0.000
Lag c_L				0.07	3.86	0.000
Constant	14.68	3.92	0.000	8.41	2.69	0.007
R^2		0.131			0.197	
Subjects		408			408	
Observations		4896			4590	

Table 4: *Random (individual) effects regressions with robust standard errors clustered at group level. Dependent variable: Followers' contributions.*

Both Endowment and Binding have a positive and significant effect on the average follower contribution. Top-up, Promise, and Returns have no significant effect in Model 1. Lab has a significant positive effect.

Model 2 adds two control variables: the leader's unconditional contribution in the current period and the leader's total contribution in the previous period. Because data for the latter variable is undefined for period 1, the number of observations is lower for Model 2 than for Model 1. Both additional controls have a positive and significant effect. Concerning the other variables, the most important change from Model 1 to Model 2 is that the effect of Endowment declines in magnitude. However, both Binding and Endowment still have a significant positive effect. Thus, it seems that these two variables' effect on the average follower contribution is partly mediated by the average leader contribution.

As explained in the previous section, our graphs indicate the presence of statistical interaction; hence, we also analyzed our data separately for treatments with even returns and for treatments with uneven returns. The results (Table 5) confirm that interaction effects are indeed present. With uneven returns ($\alpha_L = 0.64$), both Binding and Endowment have only weak positive effects that are not statistically significant at conventional levels. However, with even returns ($\alpha_L = 0.4$), each variable's effect increases by a factor of about three. They also become significant at all conventional levels. Concerning the controls, it is worth noting that Lab is no longer significant. Finally, R-squared is higher than in any of our other regressions.

on groups, does not qualitatively alter results. Inclusion of fixed-period effects (alone or in addition to random effects) does not qualitatively alter results either. Again, this holds both for individual decisions and for group averages as dependents, and for full sample analysis (Table 4) as well as for split sample analysis (Table 5).

	Even *MPCRs*			Uneven *MPCRs*		
	Coefficient	t-value	p-value	Coefficient	t-value	p-value
Binding	13.35	2.91	0.004	4.10	1.05	0.296
Endowment	19.48	3.66	0.000	5.24	1.18	0.239
Lab	2.40	0.45	0.655	14.18	3.42	0.001
Unconditional c_L	0.08	1.43	0.154	0.07	1.59	0.111
Lag c_L	0.11	3.61	0.000	0.06	2.19	0.029
Constant	11.25	3.08	0.002	20.64	4.99	0.000
R^2		0.295			0.101	
Subjects		160			152	
Observations		1800			1710	

Table 5: *Random (individual) effects regressions with robust standard errors clustered at the group level, by leader's MPCR. Treatments > 2. Dependent variable: Followers' contributions.*

In summary, our regressions support results 1 and 2. In particular, they confirm that leading by conditional commitment can enhance followers' contributions to a public good. They also confirm that this effect depends on the institutional setting. In particular, leading by conditional commitment is most likely to induce followers to contribute (more) if the leader has both credibility and ability to influence followers' payoff substantially, while the benefits from cooperation are shared evenly.

Coordination

As shown in Section 4, the leader can—by fulfilling condition (1)—create a coordination game for the followers, assuming that the leader's promise is binding. For a target of 100 ECUs in T6 it is only one way to play the positive contributions equilibrium. In treatments T4, T5, and T7 the positive contributions equilibrium requires a lower target and can be played in multiple ways. Thus, we expect leaders to set higher targets more frequently in T6 than in the other three treatments with binding promises.

Figure 2 displays the cumulative frequency of targets for the four treatments with binding promises. As can be seen, higher targets are set on average in T6 than in T4, T5, and T7. Indeed, while almost 30% of the targets in T6 are 90 ECUs or higher, only 2–12% of the targets in the three other treatments are at or above this level.

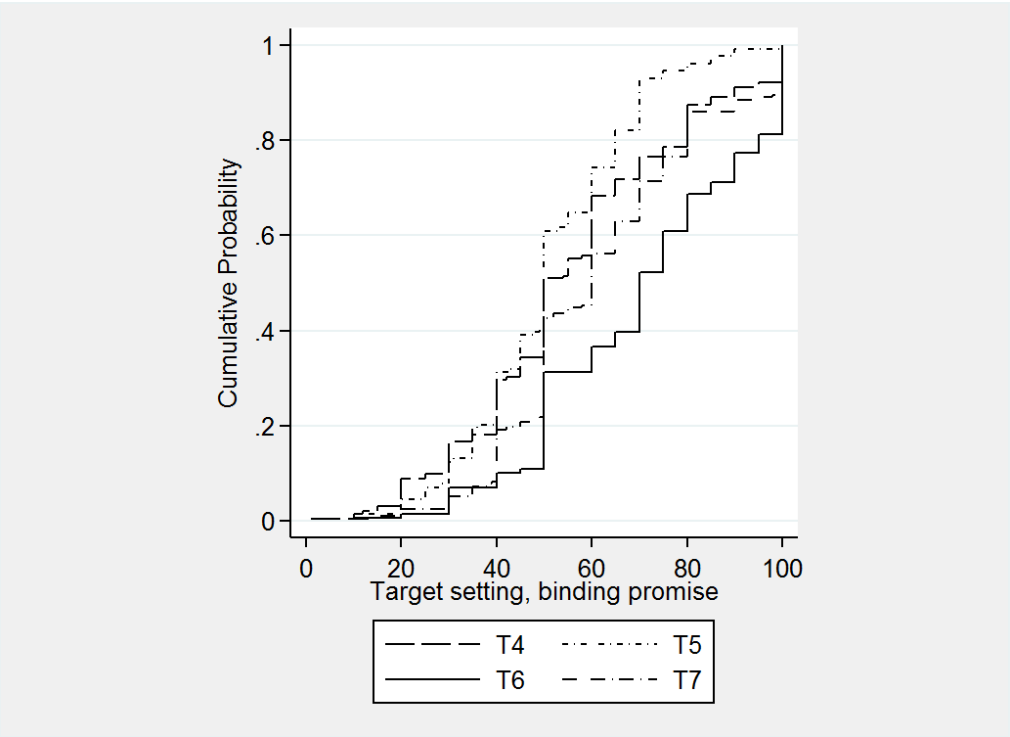

Figure 2: *Cumulative frequencies of target setting in treatments with binding promises.*

Do promises work? Figure 3 shows (a) the proportion of rounds in which condition (1) is fulfilled and (b) the proportion of rounds in which condition (1) is fulfilled *and* followers meet or exceed the leader's stated target. Treatments with non-binding promises are included for comparison.

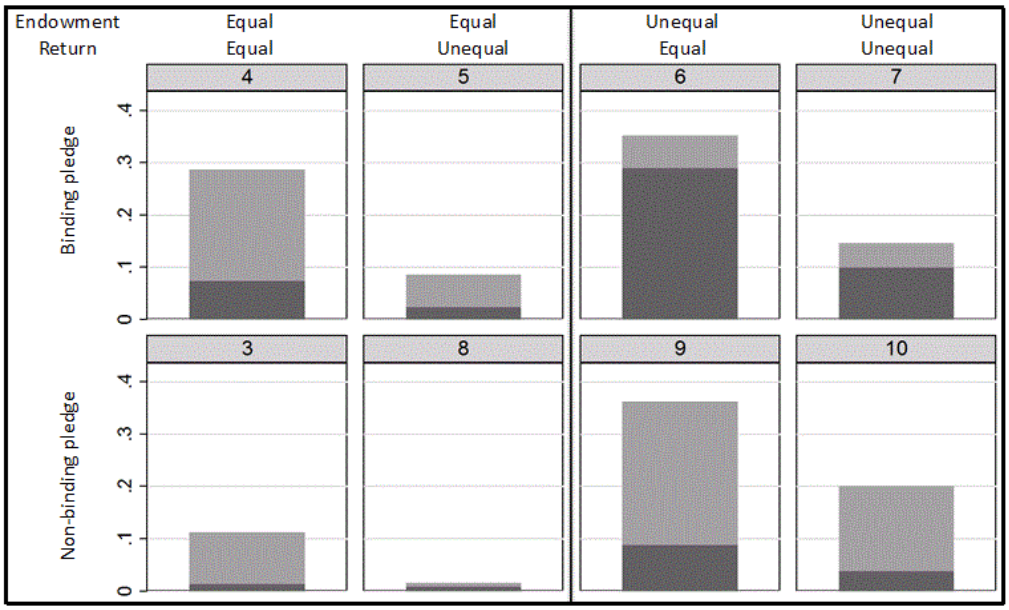

Figure 4: *Light grey bars: the proportion of rounds in which the leader fulfills condition (1). Dark grey bars: the proportion of rounds in which condition (1) is fulfilled* and *followers meet or exceed the leader's target.*

Figure 3 shows that condition 1 is met more often when returns are distributed evenly than when returns are distributed unevenly (compare T4 vs. T5, T6 vs. T7, T3 vs. T8, and T9 vs. T10). The reason is that condition 1 requires a larger promise-target ratio when returns are distributed unevenly. The figure also shows that condition 1 is met more often when the leader's endowment is large than when it is small (compare T4 vs. T6, T5 vs. T7, T3 vs. T9, and T8 vs. T10). In contrast, whether the promise is binding has no systematic effect on whether condition 1 is met. However, the fact that equilibria with positive contributions exist only when promises are binding is reflected in the followers' behavior: Binding commitments have a huge effect on whether the followers meet (or exceed) the target (compare treatments vertically). In addition, the size of the leader's endowment also affects whether the followers meet (or exceed) the target. Hence, it seems that the ability to influence followers' welfare substantially is important both because it affects the size of the leader's promise and because it affects how followers respond. In contrast, credibility is important largely because it generates equilibria with positive contributions, thereby causing followers to contribute more.

Result 3 *Leaders with a large endowment tend to promise more than do leaders with a small endowment. Followers tend to meet the leader's condition more often when the leader has a large endowment and when the leader's promise is credible.*

6 Conclusions

Our results suggest that two factors influence the prospects for leadership by conditional commitments to enhance cooperation. First, it helps if the leader's promise is credible, that is, if followers have reason to believe that fulfilling the leader's stated condition will cause the leader to actually implement the promise. Without such credibility, the followers' incentive to fulfill the leader's condition is diluted.

Second, it also helps if the leader has a large endowment, that is, if it has the ability to influence followers' welfare substantially. Indeed, unless the leader has such ability, followers cannot benefit by joining forces to fulfill the leader's condition—even when the leader's promise is credible.

Each factor's effect, however, is present only if the leader does not reap disproportionate gains from the group's collective efforts—a result that concurs with previous findings from experimental work on simultaneous-move public-good games with punishment opportunities. These previous findings show that efficient contribution norms do not easily evolve in groups where some members benefit significantly more from cooperation than others do.

Our results shed some light on why the EU's conditional commitment under the Cancun agreement largely failed to induce other major emitting countries to reciprocate. As argued by Underdal et al. (2012: 485), the EU's conditional promise to raise emissions reductions from 20% to 30% below 1990 levels may well have been credible. However, only about 11% of global emissions come from sources within the EU countries.[20] Thus, the difference between reducing EU emissions 30% and reducing them 20% corresponds to an additional reduction of global emissions of only about 1%. It is thus understandable that other major emitting countries have shown little

[20] European Commission 2014c.

interest in undertaking substantial and costly additional emissions reductions to secure such a modest global effect. In addition, these other major emitting countries face huge coordination problems in fulfilling the EU's stated condition. First, the costs of fulfilling this condition could likely be split in many different ways (however, it is hard to be sure, because the EU's stated condition is rather vague). And second, many competing norms exist concerning what is a fair division of the required mitigation burden. Thus, even if all other major emitting countries were to desire joint action to fulfill the EU's condition (which is unlikely), competing contribution norms might well undermine their ability to actually undertake such joint action. These obstacles likely apply in the INDC process as well, meaning that intrinsically conditional commitments are unlikely to facilitate ambition in the 2015 agreement. Our result that leverage matters suggests that the country most likely to influence others through conditional commitments is China, whose emissions are nearly three times those of the EU.

Our experiment suggests that the effect of leading by conditional commitments is limited even under favorable circumstances. In particular, leading by conditional commitment does not even come close to fully solving the underlying collective-action problem in any of our experimental treatments. At best, it motivates followers to contribute around half their endowment on average, so the outcome invariably remains severely suboptimal. However, although our results indicate that leadership through intrinsically conditional commitments cannot overcome the problem of climate change, they also suggest that such leadership might serve as one helpful element in a bigger package of measures. A bigger package is exactly what Victor (2011) advocates.

References

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

Barrett, S. and A. Danneberg (2012): Climate negotiations under scientific uncertainty. *Proceedings of the American Academy of Science* (PNAS) 109(43):17372–76.

Buchholz, W., C. Haslbeck, and T. Sandler (1998): When Does Partial Cooperation Pay? *Finanzarchiv* 55(1):1–20.

Chaudhuri, A. (2011): Sustaining Cooperation in Laboratory Public Goods Experiments: a Selective Survey of the Literature. *Experimental Economics* 14(1):47–83.

Cherry, T.L., S. Kroll, and J. F. Shogren (2005): The impact of endowment hetereogeneity and origin on public good contributions: evidence from the lab. *Journal of Economic Behavior & Organization* 57 (3):357–65.

Eckersley, R. (2012): Moving Forward in the Climate Negotiations: Multilateralism or Minilateralism? *Global Environmental Politics* 12(2):24–42.

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

Fischbacher, U. (2007): z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10:171–178

Fisher, J., R.M. Isaac, J.W. Schatzberg, and J.M. Walker (1995): Heterogeneous demand for public goods: Behavior in the voluntary contributions mechanism. *Public Choice* 85(3–4): 249–266.

Gächter, S., D. Nosenzo, E. Renner, and M. Sefton (2010): Who Makes a Good Leader? Cooperativeness, Optimism, and Leading-by-Example. *Economic Inquiry* 50(4): 953–967.

Güth, W., M. V. Levatia, M. Sutter, and E. van der Heijden (2007): Leading by Example With and Without Exclusion Power in Voluntary Contribution Experiments. *Journal of Public Economics* 91(5–6): 1023–1042.

Hall, D.T. (2013): *Using Role-Reversal in Laboratory Experiments.* Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2292100 (accessed 24 July, 2014).

Hoel, M. (1991): Global Environmental Problems: The Effects of Unilateral Actions Taken by One Country. *Journal of Environmental Economics and Management* 20(1): 55–70.

Holtsmark, B. (2013): International Cooperation on Climate Change: Why Is There So Little Progress? In Roger Fouquet (ed.) *Handbook on Energy and Climate Change.* Cheltenham, UK: Edward Elgar.

Isaac, R. M. and J. Walker (1988): Communication and free-riding behavior: The voluntary contribution mechanism. *Economic Enquiry* 26(2): 585–608.

Konrad, K. A. and M. Thum (2014): Climate Policy Negotiations with Incomplete Information. *Economica* 81(322): 244–256.

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

Levati, M. V., M. Sutter and E. van der Heijden (2007: Leading by Example in a Public Goods Experiment with Heterogeneity and Incomplete Information," *Journal of Conflict Resolution* 51(5): 793–818.

Maltais, A. (2014): Failing International Climate Politics and the Fairness of Going First. *Political Studies* 62(3): 618–633.

McEvoy, D. M. (2010: Not it: Opting out of Voluntary Coalitions that Provide a Public Good," *Public Choice* 142(1): 9–23.

McEvoy, D. M., T. L. Cherry, and J. K. Stranlund (2011): *The Endogenous Formation of Coalitions to Provide Public Goods: Theory and Experimental Evidence.* Department of Economics, Appalachian State University. Working paper.

McLean, E and R. Stone (2012): The Kyoto Protocol: Two-Level Bargaining and European Integration. *International Studies Quarterly* 56(1): 99–113.

Ostrom, E. (2000): Collective Action and the Evolution of Social Norms. *Journal of Economic Perspectives* 14(3): 137–158.

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

Ott, H. E. and S. Oberthür (1999): *Breaking the Impasse. Forging an EU Leadership Initiative on Climate Change.* Policy paper. Berlin: Heinrich Böll Stiftung.

Reuben, E. and R. Riedl (2013): Enforcement of Contribution Norms in Public Good Games with Heterogeneous Populations. *Games and Economic Behavior* 77(1): 122–137.

Rivas, F. and M. Sutter (2011): The Benefits of Voluntary Leadership in Experimental Public Goods Games. *Economics Letters* 112(2): 176–8.

Schelling, T. C. (1960): *The Strategy of Conflict.* Ann Arbor: Harvard University Press.

Selten, R. (1967): Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In: H. Sauermann (ed.) *Beiträge zur experimentellen Wirtschaftsforschung.* Tübingen: Mohr.

Skodvin, T. and S. Andresen (2006): Leadership Revisited," *Global Environmental Politics* 6(3): 13–27.

UNFCCC (2015): https://unfccc.int/documentation/documents/advanced_search/items/6911.php?priref=600008407 (accessed 2 June 2015)

UNFCCC (2011a): http://unfccc.int/resource/docs/2010/cop16/eng/07a01.pdf (accessed 2 June 2015)

UNFCCC (2011b): http://unfccc.int/resource/docs/2011/sb/eng/inf01r01.pdf (accessed 2 June 2015)

Underdal, A., J. Hovi, S. Kallbekken, and T. Skodvin (2012): Can Conditional Commitments Break the Climate Change Negotiations Deadlock?" *International Political Science Review* 33(4): 475–493.

Victor, D. (2011): *Global Warming Gridlock. Creating More Effective Strategies for Protecting the Planet.* Cambridge, MIT: Cambridge University Press.

Weischer, L., J. Morgan and M. Patel (2012): Climate Clubs: Can Small Groups of Countries Make a Big Difference in Addressing Climate Change?" *Review of European Community and International Environmental Law* (RECIEL) 21(3): 177–192.

Weimann, J., J. Brosig-Koch, H. Hennig-Schmidt, C. Keser and C. Stahr (2012): *Public-good experiments with large groups.* Working-paper: Faculty of Economics, Georg-August-Universität Göttingen.