



# Working Papers

[www.cesifo.org/wp](http://www.cesifo.org/wp)

## To Commit or not to Commit? An Experimental Investigation of Pre-Commitments in Bargaining Situations with Asymmetric Information

Sönke Hoffmann  
Benedikt Mihm  
Joachim Weimann

CESIFO WORKING PAPER NO. 4835  
CATEGORY 13: BEHAVIOURAL ECONOMICS  
JUNE 2014

*An electronic version of the paper may be downloaded*

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: [www.CESifo-group.org/wp](http://www.CESifo-group.org/wp)

# To Commit or not to Commit? An Experimental Investigation of Pre-Commitments in Bargaining Situations with Asymmetric Information

## Abstract

In a recent paper Konrad and Thum (2014) present a model that shows that unilateral pre-commitment reduces the likelihood of agreement in bilateral negotiations over the provision of a public good when parties have private information over their contribution costs. We test the model in a laboratory experiment paying particular attention to how behavioral motivations other than payoff-maximization affect the strength of the model's result. We find that the result is no longer statistically significant when we allow for non-payoff-maximizing behavior at each stage of the game. Introducing communication has an interesting effect as it influences different forms of non-payoff-maximizing behavior asymmetrically and leads to the model's result again becoming significant. All in all, we find strong experimental support for Konrad and Thum's model even though we observe considerable amounts of non-payoff-maximizing behavior that is not accounted for in the original model.

JEL-Code: C900.

Keywords: pre-commitments, bargaining, asymmetric information, social preferences.

*Sönke Hoffmann*  
*Faculty of Economics and Management*  
*Otto-von-Guericke University*  
*Magdeburg / Germany*  
*soenke.hoffmann@ovgu.de*

*Benedikt Mihm*  
*Faculty of Economics and Management*  
*Otto-von-Guericke University*  
*Magdeburg / Germany*  
*ben.miehm@ovgu.de*

*Joachim Weimann\**  
*Faculty of Economics and Management*  
*Otto-von-Guericke University*  
*Magdeburg / Germany*  
*joachim.weimann@ww.uni-magdeburg.de*

\*corresponding author

April 9, 2014

## 1. Introduction

Bargaining over the private provision of public goods may lead to inefficient outcomes if parties have incomplete information or contracts are not enforceable. The literature has largely concentrated on the enforcement problem. Work on international negotiations on environmental regulations, for example, has paid particular attention to the enforcement problem because no common institution exists in this setting and the enforcement of contracts is thus difficult (Carraro and Siniscalco, 1993; Barrett, 1998).

In a recent paper Konrad and Thum (2014) focus instead on the problems that arise in a bargaining environment with asymmetric information. Their model (referred to as “KT-model” henceforth) assumes the enforcement problem is resolved and examines bargaining over contributions to a public good when parties are privately informed about their cost of provision.

Under asymmetric information bargaining outcomes will generally be inefficient as negotiations can break down with a positive probability even when mutually beneficial agreements are possible (Meyerson and Satterthwaite, 1983). It is well known that in markets for private goods the inefficiency disappears as the number of traders increases and the market becomes large (Gresik and Satterthwaite, 1989). However, Rob (1989) showed that even this asymptotic efficiency does not hold for public goods and thus under asymmetric information negotiations over the private provision of a public good are unlikely to ever achieve an efficient solution.

The question remains, however, how large the inefficiencies will be and under what kind of negotiation rules the likelihood of negotiation breakdown, and thus the inefficiency, can be minimized. In particular, it is unclear if prior commitments by one party have a positive influence on the prospects for achieving more efficient outcomes. The KT-model makes an important contribution to the literature on the private provision of public goods by investigating this issue in a non-cooperative game setting.

The question on the role of prior commitments is highly relevant. The EU,

for example, seems to view pre-committing to environmental damage prevention as an act that sets a good example for others and that will motivate others to follow suit. The KT-model, however, states the exact opposite. Comparing the equilibria of two sequential bargaining games – one with commitment and one without – the authors show that the probability for successful cooperation is strictly lower when one party has contributed to the public good before bargaining takes place. This result obviously has strong political implications.

Our paper is an experimental investigation of the findings of the KT-model. In addition to a direct experimental verification of the model our experiment focuses on the potential for the bargaining situation modeled by Konrad and Thum to be influenced by various motives that deviate from payoff-maximization and which could thus affect the results of the model. Inequality aversion, for example, might prevent players from payoff-maximizing if payoff differences are sufficiently large (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Direct and indirect reciprocity (Falk and Fischbacher, 2006; Nowak and Siegmund, 2005) may also be a factor in the presence of pre-commitments.

Results from a world-wide survey of people involved in international climate policy indicate that fairness and equity considerations can play a significant role in climate negotiations (Lange et al., 2007, 2010). Since the KT-model's results are particularly relevant for climate negotiations it is thus important to investigate how the model performs in a bargaining environment in which real subjects may harbor such behavioral motivations. A laboratory setting is the ideal venue to explore this extension as the experiment can control for the amount of freedom subjects have to deviate from payoff-maximization.

In order to find out how behavioral motivations other than payoff-maximization affect the results of the KT-model the experiment is designed to be carried out in three settings. Each of the three settings has one treatment with pre-commitment and one without. The first setting is intended to be a direct assessment of the KT-model as it most closely follows the basic assumptions of the original theory, i.e., payoff-maximizing behavior and common knowledge. Technically, one subgame of the KT-model (standard prisoners' dilemma) is re-

placed by the corresponding Nash payoffs and thus players are forced to behave in a payoff-maximizing way in the final stage of the game. In our experiment we found that in this reference setting cooperation took place twice as often in the treatment without pre-commitment when compared to the treatment with pre-commitment (referred to as “cooperation gap” henceforth).

In the second setting the entire prisoners’ dilemma is re-introduced to ascertain whether the KT-model is affected by giving subjects additional room to behave in non-payoff-maximizing ways, and if so, whether the cooperation gap persists. We found that the gap did persist in our experiment but became considerably smaller.

In the third setting the KT-model is pushed even further away from its original assumptions through the introduction of pre-play communication between the bargaining parties. There are two motivations for this extension. First, the experimental literature on the provision of public goods has shown that communication between subjects increases the level of cooperation even if communication is cheap talk (Brosig et al., 2003; Valley et al., 1998). It is still unclear, however, what effect communication has in environments with or without pre-commitment. Second, it is an artificial assumption that bargaining over the provision of public goods takes place without communication between the parties involved. It is thus important for the external validity of the KT-model to check whether or not it is “communication proof”. In fact, in our experiment we observed that with communication there was a strong increase in success rates in both the pre-commitment and no pre-commitment treatments but at the same time the cooperation gap again opened significantly.

The remainder of this paper is structured as follows. The next section outlines the KT-model as it was implemented in our experiment. In section three we specify the experimental procedure. Section four contains our main results, and in the final section five we discuss our findings.

## 2. The KT-Model

The KT-Model encompasses two variants of a sequential bargaining game, one with pre-commitment and one without. We start with the more general version without pre-commitment.

Two players  $i \in \{A, B\}$  negotiate over the provision of a public good  $e = e_A + e_B$ , where  $e_A$  and  $e_B$  denote the contribution of player  $A$  and  $B$  respectively. Both players can either make a contribution ( $e_i = 10$ ) or not ( $e_i = 0$ ). If player  $i$  decides to contribute, his cost of contribution is  $10 + c_i$  with  $c_i \in \{1, 2, \dots, 9\}$ . The cost parameter  $c_i$  is private information of player  $i$  and is randomly drawn from a uniform distribution. In the bargaining process, player  $A$  can offer a transfer  $t \in \{-10, -9, \dots, 9, 10\}$  to player  $B$ . If  $t > 0$  the transfer goes from  $A$  to  $B$  which means that  $A$  pays a price to  $B$ , if  $t < 0$  the transfer is a price  $B$  pays to  $A$ .<sup>1</sup>

The overall bargaining structure is characterized by a “take it or leave it” offer similar to the classic ultimatum game: Player  $A$  proposes a transfer to  $B$  which  $B$  can accept or reject. If  $B$  accepts then both players become obliged to contribute to the public good ( $e_i = 10$ ). If the offer is rejected no transfer is paid and both players decide over their contributions independently. In this case both players are in a prisoners-dilemma and choosing not to contribute is their dominant strategy. The payoffs of the players can be written as

$$\pi_A = e_B - c_A e_A - t \quad \text{and} \quad \pi_B = e_A - c_B e_B + t. \quad (1)$$

Figure 1 visualizes the sequential structure of the game without pre-commitment. This version of the model is contrasted with a version in which  $A$  makes a commitment before the game starts. Technically, this pre-commitment is modeled by fixing  $e_A = 10$  throughout the whole game, which removes strat-

---

<sup>1</sup>Note that the original model not only applies to a uniform distribution but to any distribution that satisfies a non-negative inverse hazard rate. Furthermore, Konrad and Thum use continuous cost and transfer values, whereas we use integer values scaled by factor ten in order to ensure experimental compatibility.

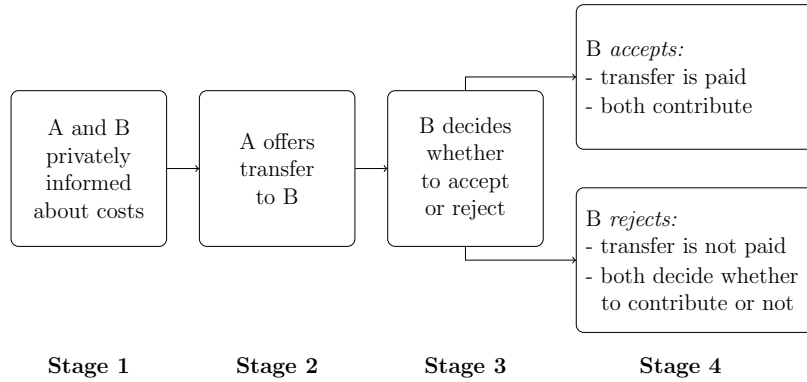


Figure 1: Sequential structure of the game without pre-commitment.

egy  $e_A = 0$  from the prisoners' dilemma in the last stage. Thus, player  $A$  no longer decides about his contribution and this is common knowledge.

Under the assumption of payoff-maximization the KT-model has the following two results.

**Result 1.**

*The probability of the bargain being successful is strictly larger in the game without pre-commitment than it is in the game with pre-commitment.*

**Result 2.**

*The unique perfect Bayesian equilibrium transfers are bound by  $(-10, 0]$  in the game without pre-commitment and by  $(0, 10)$  in the game with pre-commitment. Specifically, under the conditions implemented in the experiment the equilibrium transfers are given by  $t_{nPC}^* = -c_A/2$  in the game without pre-commitment and  $t_{PC}^* = 5$  in the game with pre-commitment.*

The intuition behind these results is as follows. If player  $A$  does not pre-commit before bargaining takes place then his gain from reaching an agreement is greater. To keep the chances of getting this gain realized  $A$  has to bargain less aggressively which enhances the likelihood of cooperation relative to the game with pre-commitment.

Furhtermore, if player  $A$  does not pre-commit then he can sell his willingness to cooperate to the  $B$ -player.  $A$  thus demands a price for cooperation and we have  $t \leq 0$ . If  $A$  pre-commits, however, he needs to offer  $B$  something to make him cooperate and we have  $t > 0$ .

### 3. The Experiment

#### 3.1. Background

The experiment is designed to detect the influence of behavioral motivations other than payoff-maximization. Generally,  $B$ -players can deviate from payoff-maximizing behavior in two ways. They may either accept the offer although rejecting would provide a higher payoff, or reject the offer although accepting would provide a higher payoff.

A non-payoff-maximizing rejection may be the result of negative reciprocity. Negative reciprocity refers to a situation in which people take revenge even in interactions with complete strangers, and even if it is costly for them and yields neither present nor future material rewards (Fehr and Gächter, 2000). If, for example,  $A$  offers a transfer that  $B$  expected to be higher, then  $B$  may see a bad intention behind this action. Consequently,  $B$  could reject the offer to punish  $A$  even if this would also reduce his own payoff. Another potential cause for non-payoff-maximizing rejections is inequality aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). In this case,  $B$  rejects the offer at the cost of lower own payoffs because this choice reduces the inequality between both players' payoffs.

A non-payoff-maximizing acceptance can be explained by positive (indirect) reciprocity. In particular, when  $A$  has to pre-commit a non-payoff-maximizing acceptance may be driven by upstream reciprocity, which is a form of indirect reciprocity of the type "somebody else helped me and I help you" (Nowak and Roch, 2006). Since the introduction of a fixed pre-commitment puts the  $B$ -players in a dominant position it may trigger them to reward the  $A$ -players at a cost to themselves even though the pre-commitment was exogenously imposed



on  $A$ .<sup>2</sup> Another reason why  $B$  may accept although this reduces own payoff is a general willingness to cooperate. Player  $B$  may simply prefer to set up the common good instead of acting in pure self-interest.

For the  $A$ -players, not offering high enough transfers in the treatments with pre-commitment may also be motivated by inequality aversion. With pre-commitments the  $A$ -players are already in a disadvantageous position relative to the  $B$ -players so they may be unwilling to make high offers as this would further increase the inequality of the payoffs.

With these behavioral motivations in mind, the experiment is carried out in three settings that give subjects varying amounts of freedom to deviate from payoff-maximization. In all settings we have one treatment with and one without pre-commitment.

- **Setting 1** is the reference setting designed to control for non-payoff-maximizing behavior. If an offer is rejected both players' contributions are fixed to the dominant strategy  $e_i = 0$  to force them to play the prisoners' dilemma's Nash equilibrium. The two treatments in this setting are labeled `PC_nD_nC` (pre-commitment, no decision, no communication) and `nPC_nD_nC` (no pre-commitment, no decision, no communication).
- **Setting 2** gives both players more room for non-payoff-maximizing behavior because at the last stage of the game they are free to choose whether to contribute to the public good or not. We conjecture that having a choice in the final stage of the game may have a feedback effect on the previous stages. If subjects are completely rational we should observe no such feedback and we should thus observe no difference between Setting 1 and 2. The two treatments in Setting 2 are labeled `PC_D_nC` (pre-commitment,

---

<sup>2</sup>Note that upstream reciprocity originated from Evolutionary Game Theory which analyzes the evolution of populations given many repetitions of the game. The behavioral pattern we observe here is closest to the character of upstream reciprocity, even though our games were played one-shot.

Table 1: Overview of all treatments played.

Setting	Name	pre-play commitment	post-rejection choice	pre-play communication
1	nPC_nD_nC	No	No	No
	PC_nD_nC	Yes	No	No
2	nPC_D_nC	No	Yes	No
	PC_D_nC	Yes	Yes	No
3	nPC_D_C	No	Yes	Yes
	PC_D_C	Yes	Yes	Yes

decision, no communication) and nPC\_D\_nC (no pre-commitment, decision, no communication).

- **Setting 3** is identical to Setting 2 but includes a three minute pre-play chat using a chat-box integrated into the user interface. The opportunity for communication occurs in stage one of the game after the subjects learn their cost of provision. Written content is essentially unrestricted but any information that reveals a player’s identity is prohibited. The conjecture is that subjects will behave more cooperatively in the communication treatments as has been observed in various experiments on public good provision (Brosig et al., 2003; Valley et al., 1998). The two treatments in this setting are labeled PC\_D\_C (pre-commitment, decision, communication) and nPC\_D\_C (no pre-commitment, decision, communication).

Table 1 gives an overview of the three pairs of treatments played in the three settings.

### 3.2. Experimental setup

We used a between subject design such that each subject participated in one session and in each session the subjects were exposed to just one of the

treatments. In all three settings we had six sessions per treatment and ten subjects participated in each session. At the beginning of a session subjects were randomly selected into the role of either an *A*- or a *B*-player. These roles were fixed throughout the session. Within each session every *A*-player was paired once with every *B*-player (“round robin”).<sup>3</sup> Thus, every subject played the game five times and in total we had 150 observations per treatment (5 pairs  $\times$  5 rounds  $\times$  6 sessions). Since each subject faced any other subject at most once we take all 150 observations to be independent. In total 360 subjects participated in the experiment (3 settings  $\times$  2 treatments  $\times$  6 sessions  $\times$  10 subjects).

After the roles had been fixed the subjects were given written instructions and sufficient time to read them.<sup>4</sup> Subjects were informed that their possible cost values ranged from 1 to 9 and were randomly selected with the same probability. Once drawn, we used the same values in all sessions to make sessions comparable. The experiment started with three practice rounds in which the subjects played against the computer. All participants were informed that the computer played payoff-maximizing strategies throughout the practice rounds and that these rounds were not payoff relevant. In the first two practice rounds subjects learned the computer’s cost as well as their own. The third round simulated the actual game as each player was informed only about their own cost of provision.

At the start of the actual game the subjects were given an initial endowment to ensure that it was not possible for them to make a loss. Subjects were informed that they received an endowment that covered losses but not how high the endowment actually was.<sup>5</sup> Subjects were not paid a show up fee on

---

<sup>3</sup>The round robin structure was used so as to allow five observations to be generated per subject without losing the one shot character of the game. As Kamecke (1997) has shown this is most likely to be the case using round robin.

<sup>4</sup>See example of instruction in Appendix D.

<sup>5</sup>All subjects were given an endowment of five euro. The *A*-players in the pre-commitment treatments were given an additional five euro to compensate them for being in a disadvantaged

top of this endowment. During the game subjects had access to on-screen tables that provided information about relevant payoffs.<sup>6</sup> These tables included their own payoffs conditional on their and the other players possible decisions, as well as the payoffs of the other player conditioned on the other player’s possible costs. The purpose of this information was to make the game easier to follow and to minimize calculation effort.

The experiment was carried out at the experimental laboratory at the University of Magdeburg, Germany (MaXLab) and was programmed using z-Tree (Fischbacher, 2007). The sessions lasted on average 40 : 16 minutes. At the end of the experiment the payoffs of all five games were paid and the average earnings of the subjects were 10.35 euro.

## 4. Results

### 4.1. Verification of Results 1 and 2

In this section we address the question whether the theoretical predictions in Result 1 and 2 are compatible with the experimental data of our reference setting (Setting 1), and if so, whether these results still hold given a post-rejection decision (Setting 2) and communication (Setting 3). Table 2 summarizes the (aggregated) experimental data of each treatment. Player  $A$ ’s behavior is captured by the average transfer  $\bar{t}$  and the average deviation  $\Delta t^*$  from the predicted equilibrium value. The basic behavior of player  $B$  is described by  $a$ , the total number of acceptances and  $r$ , the total number of rejections. Dividing  $a$  by the total number of decisions (150 in each treatment) gives the total acceptance rate that is plotted in Figure 2 over all three settings.

Result 1 predicts a higher probability of successful negotiation if there is no pre-commitment. In the first setting of our experiment we observed that the acceptance rate was 72% in the nPC-condition and 36% in the PC-condition (cf.

---

position. It is due to this difference that we did not inform the subjects about the value of their endowments.

<sup>6</sup>See screenshots in Appendix E.

Table 2: Behavioral results of all treatments. (1) Behavior of  $A$ :  $\emptyset t$ : Average transfer offered (Lab Dollars),  $\Delta t^*$ : Average difference between  $t$  and the theoretical equilibrium value  $t^*$  (Lab Dollars), *median t*: Median values of the transfers. (2) Expected behavior of  $B$  (absolute numbers):  $a^{hyp}$ : hypothetical acceptance cases (payoff from accepting was strictly higher),  $r^{hyp}$ : hypothetical rejection cases ((expected) payoff from rejecting was strictly higher),  $i^{hyp}$ : hypothetical indifference cases (accepting and rejecting had same payoffs) (3) Actual behavior of  $B$ :  $a$ : total acceptances,  $a_{pmx}$ : acceptances when accepting was the payoff-maximizing choice,  $a_{npx}$ : acceptances when accepting was not the payoff-maximizing choice,  $a_{ind}$ : acceptances when accepting and rejecting generated equal payoffs (indifference). All rejection variables have the equivalent interpretation given that  $B$  rejected the offer. Column “totals” sums up the respective acceptance and rejections such that  $\omega = a + r$ , respectively

Treatm.	Behavior of players																
	A			B's exp. beh.			B accepted			B rejected			totals				
	$\emptyset t$	$\Delta t^*$	<i>median t</i>	$a_{hyp}$	$r_{hyp}$	$i_{hyp}$	$a$	$a_{pmx}$	$a_{npx}$	$a_{ind}$	$r$	$r_{pmx}$	$r_{npx}$	$r_{ind}$	$\omega_{pmx}$	$\omega_{npx}$	$\omega_{ind}$
nPC_nD_nC	-1.69	0.71	-2.00	129	9	12	108	104	0	4	42	9	25	8	113	25	12
PC_nD_nC	0.77	-4.23	2.00	38	102	10	54	35	11	8	96	91	3	2	126	14	10
nPC_D_nC	-1.88	0.52	-1.00	122	20	8	82	81	1	0	68	19	41	8	100	42	8
PC_D_nC	1.98	-3.02	3.00	49	88	13	70	49	10	11	80	78	0	2	127	10	13
nPC_D_C	0.22	2.62	0.00	147	2	1	137	136	1	0	13	1	11	1	137	12	1
PC_D_C	2.84	-2.16	3.00	52	76	22	109	50	39	20	41	37	2	2	87	41	22

Figure 2: Variations of the cooperation gap in three experimental settings.

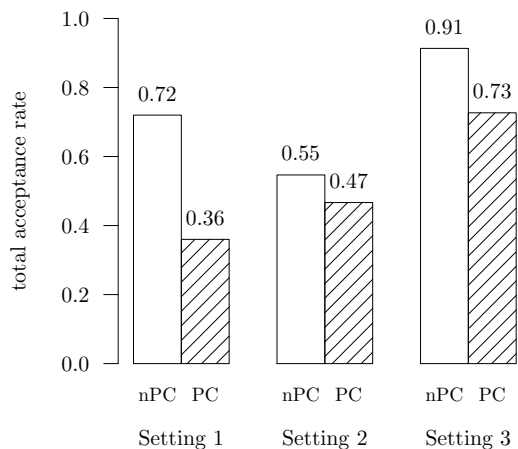


Fig. 2, Setting 1). Clearly, this difference is statistically significant ( $p$ -value  $< 0.001$ ,  $\chi^2$ -test). Allowing subjects a post-rejection decision in the prisoners dilemma (Setting 2) led to the cooperation gap becoming much smaller (55% vs. 47%) and no longer being statistically significant ( $p$ -value = 0.204,  $\chi^2$ -test). Result 1 is, nevertheless, confirmed qualitatively. Finally, adding pre-play communication (Setting 3) increased the agreement rates with and without pre-commitment. However, the influence of communication was much stronger without pre-commitment such that the cooperation gap once again opened widely (91% vs. 73%) and the difference between the rates once again becomes statistically significant ( $p$ -value  $< 0.001$ ,  $\chi^2$ -test).

Without pre-commitment Result 2 predicts equilibrium transfers  $t_{nPC}^*(c_A) = -0.5c_A$ . Using our experimental data to fit the linear model  $t_{nPC} = \alpha + \beta c_A$ , we find that neither  $H_0 : \alpha = 0$  nor  $H_0 : \beta = -0.5$  can be rejected in Setting 1 and 2. In Setting 3 only the slope parameter differs significantly from the predicted value. A simultaneous  $F$ -test on both parameters identifies the estimated model in Setting 2 as not statistically

different from the theoretical prediction.<sup>7</sup>

The introduction of pre-commitment had two effects on transfer behavior. On the one hand, we no longer observed any evident relationship between  $t$  and  $c_A$ . At all cost levels the transfers spread over their maximum range with correlation coefficients between 0.01 and 0.1. Qualitatively, this is compatible with the model which predicts a constant transfer value ( $t_{PC}^* = 5$ ) that is completely unrelated to the cost level  $c_A$ . On the other hand, however, the average observed transfer  $\bar{t} = 0.77$  is considerably smaller than the theoretically predicted value 5.

All in all, we see strong experimental evidence in support of the model, even under conditions that may deviate from the original KT-model.

#### 4.2. Behavioral analysis

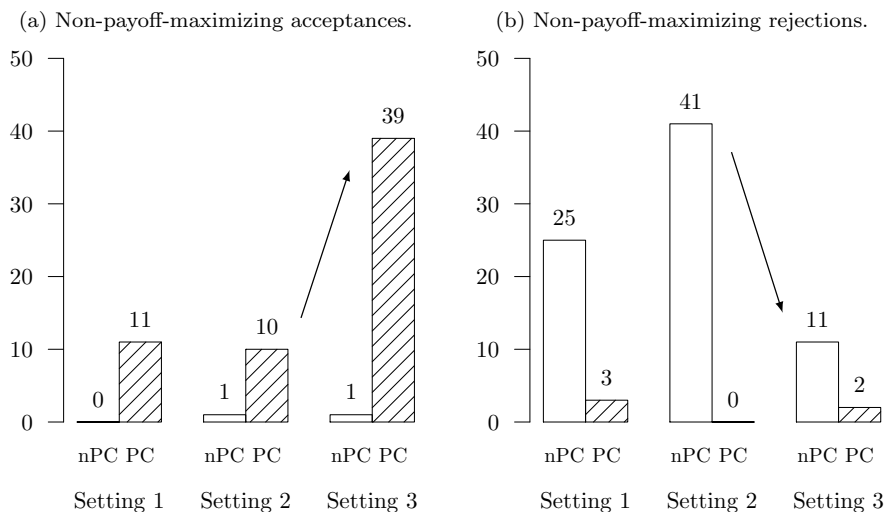
We now address the question of how the interplay between different behavioral motivations resulted in such strong support for the KT-model's main predictions. The behavioral analysis focuses on non-payoff-maximizing behavior and requires a number of refinements of the measures  $a$  and  $r$ . The variables in columns 4 – 6 of Table 2 represent the number of cases for which the transfer-cost constellation made accepting generate a strictly higher payoff than rejecting ( $a_{hyp}$ ), rejecting generate a strictly higher payoff than accepting ( $r_{hyp}$ ), and both decisions generate an equal payoffs ( $i_{hyp}$ ).  $a_{hyp}$  is split up into  $a_{pmx}$ , the number of times  $B$ -players indeed accepted a profitable offer and  $r_{npmx}$ , the number of times  $B$ -players rejected even though accepting was the payoff-maximizing choice. Similarly, the total number of acceptances  $a$  is the sum of  $a_{pmx}$  and  $a_{npmx}$ , where the latter is the number of times  $B$ -players who accepted when rejecting would have provided a higher payoff. Equivalent decompositions hold for  $r_{hyp}$  and  $r$ .

Figure 3 shows the number of non-payoff-maximizing acceptances  $a_{npmx}$  (Fig. 3a) and rejections  $r_{npmx}$  (Fig. 3b) of the  $B$ -players in all three settings.

---

<sup>7</sup>A more detailed exposition of the regression results can be found in Appendix A.

Figure 3: Deviations from payoff-maximization over all treatments. Values above the bars represent the absolute number of  $B$ -players. The arrows indicate the effect of communication.



Without pre-commitment we did not find any indication of positive reciprocity or cooperative behavior in the  $B$ -players' behavior. Acceptances were clearly payoff driven as the white bars in Figure 3a show no deviations from payoff-maximization in the nPC-treatments. This is quite different when it comes to the rejections. A considerable number of transfers were rejected although accepting would have provided a higher payoff (cf. white bars in Fig 3b). This indicates the presence of negative reciprocity or inequality aversion among  $B$ -players.<sup>8</sup>

With pre-commitment the pattern was completely reversed:  $B$ 's rejection behavior was in line with payoff-maximization but the acceptance behavior was not.  $A$ 's disadvantageous position (relative to  $B$ 's) caused by the pre-commitment seems to have triggered  $B$ 's willingness to accept offers even though

<sup>8</sup>As the transfer offered by the  $A$ -player is essentially an ultimatum offer it is not a surprising that  $B$ -players were willing to reject offers which would make them better off but were perceived as unfair. This kind of behavior is well known from ultimatum game experiments (Güth et al., 1982).



rejecting would have provided a higher payoff.

It is important to note that the non-payoff-maximizing behavior both in regards to acceptances and rejections work against Result 1 of the KT-model. The non-payoff-maximizing acceptances in the PC-treatments increased the amount of agreements in the games with pre-commitment and the non-payoff-maximizing rejections in the nPC-treatments reduced the amount of agreements in the games without pre-commitment. Given the overall characterization of the behavior of the  $B$ -players it is surprising that the predictions of the KT-model are nevertheless confirmed by our findings. To see why this is the case, we have to look more closely at the three settings and we have to take the behavior of the  $A$ -players into account.

#### *4.2.1. Setting 1: No last stage decision, no communication*

In our reference setting we observed the general behavioral pattern outlined above: Without pre-commitment acceptances were payoff maximizing and rejections were not (negative reciprocity), whereas with pre-commitment rejections were payoff-maximizing and acceptances were not (positive reciprocity). As negative reciprocity reduces the number acceptances and positive reciprocity increases it, both observed effects work against the Result 1 of the KT-model.

The reason that there were nevertheless significantly less agreements in the PC-treatment than in the nPC-treatment in this setting can be attributed to the fact the offers made by the  $A$ -players were not high enough in the PC-treatment. Only 38 of the 150 offered transfers in the PC-treatment would have made the  $B$ -player better off by acceptance than by rejection, compared to 129 of the 150 offers for which this would have been the case in the nPC-treatment.

The KT-model predicts a transfer payment in the PC game of  $t_{PC}^* = 5$  but the average transfer payments in the PC-treatment was only 0.77. This transfer behavior is consistent with inequality aversion because the  $A$ -players were already required to pre-commit to the provision of the public good. If they additionally paid a transfer of 5 then their own payoff would have been  $\pi_A = 5 - c_A$  and the payoff of the  $B$ -player would have been  $\pi_B = 15 - c_B$ . It

is plausible that the  $A$ -players considered this inequality in payoffs too extreme and were thus not willing to pay the required price to reach an agreement. As a result, the agreement rate fell steeply enough that the difference in the number of agreements reached in the PC-treatment and the nPC-treatment was so pronounced.

#### 4.2.2. *Setting 2: Last stage decision, no communication*

Under the PC-condition the behavior of the  $B$ -players did not change much between Setting 1 and Setting 2. Figure 3 shows that the introduction of a contribution choice in the prisoners' dilemma made non-payoff-maximizing rejections change from 3 to 0 and the non-payoff-maximizing acceptances fall from 11 to 10. Put differently, we still observed payoff-maximizing rejection behavior and moderate positive reciprocity or cooperative behavior from the  $B$ -players. The transfer behavior of the  $A$ -players did not change much either, as the  $a_{hyp}$  values in Setting 1 and 2 are quite similar.<sup>9</sup>

The key difference between the two settings can be found in the nPC-condition. In Setting 1 the  $B$ -players rejected  $r_{npmx} = 25$  out of  $a_{hyp} = 129$  (= 19%) advantageous offers made by  $A$  and in Setting 2 this rate was 41 out of 122 (= 34%). Thus, the tendency towards non-payoff-maximizing rejections by the  $B$ -players increased when both players were free to choose their contribution in the prisoners' dilemma. This, in turn, made the number of agreements fall and the cooperation gap close.

This raises the question as to why  $B$ -players rejected offers even though the resulting payoff in the non-cooperative solution of the prisoners' dilemma was strictly lower than the safe payoff from accepting. In the treatment in Setting 2 without pre-commitment this type of behavior is only reasonable if the  $B$ -players expected that the  $A$ -players would not play their dominant strategy in the prisoners' dilemma game. If all players expected every other player to choose

---

<sup>9</sup>This behavior is not too surprising as  $A$ 's decision space is exactly the same in both settings when he has to pre-commit.

the non-cooperative strategy in the prisoners' dilemma then there should be no difference in the  $r_{npmx}/a_{hyp}$  values between Setting 1 and 2. But as we did observe a difference it can be attributed to  $B$ 's expectation to exploit the other player in the final stage.

#### 4.2.3. Setting 3: Last stage decision, communication

Adding pre-play communication in Setting 3 neither affected non-payoff-maximizing acceptances under the nPC-condition nor did it affect non-payoff-maximizing rejections under the PC-condition.<sup>10</sup> With communicating  $B$ -players still did not reward  $A$ -players at cost to themselves when  $A$  was not disadvantaged by pre-commitment, and  $B$ -players still did not punish  $A$ -players at cost to themselves when  $A$  was disadvantaged by pre-commitment. However, as the arrows in Figure 3 show communication did increase positive reciprocity under PC ( $a_{npmx}$  increased from 10 to 39) and reduced negative reciprocity under nPC ( $r_{npmx}$  falls from 41 to 11) by nearly the same amount. Therefore, communication influences  $B$  towards cooperation in two different ways: On the one hand it led  $B$  to share in the disadvantage of  $A$ 's pre-commitment and accept non-profitable transfers ( $a_{npmx}$  increases), on the other hand it neutralized the expectations  $B$ 's had of exploiting  $A$  that were present without communication in the nPC-condition ( $r_{npmx}$  falls). The first effect works against Result 1 of the KT-model, the second one supports it.

If communication had no other effect we should have observed that the cooperation gap remained unchanged. The reason why the cooperation gap opened again in Setting 3 is that communication also affected the generosity of the  $A$ -players. Average transfers  $\varnothing t$  as well as the number of offers  $a_{hyp}$  that a payoff-maximizing actor would accept take their highest values in Setting 3.<sup>11</sup>

---

<sup>10</sup>See Appendix C for a detailed discussion of how communication was used by the subjects. It is worth noting here that the opportunity to communication was utilized extensively in both treatments although more often in the treatment without pre-commitment.

<sup>11</sup>Note that this effect of communication was so strong that even the  $\varnothing t$  in the PC-condition became positive, which is difficult to explain in the context of the KT-Model.

This made the number of payoff-maximizing acceptances rise by  $\Delta a_{pmx} = 55$  and the number of payoff-maximizing rejections change by  $\Delta r_{pmx} = -18$  under the nPC-condition. In total this effect is stronger than under the PC-condition ( $\Delta a_{pmx} = 1$  and  $\Delta r_{pmx} = -41$ ), supporting Result 1 of the model.

#### 4.3. Estimation of treatment effects

As a final piece of analysis we use our experimental data to estimate the logistic regression model

$$\Pr(Acc_B = 1) = \Lambda(\alpha + \beta_1 t + \beta_2 c_B + \gamma_1 D_1 + \dots + \gamma_5 D_5) \quad (2)$$

in which  $Acc_B$  (“Did Player B accept?” Yes = 1, No = 0) represents the binary outcome variable and  $\Lambda(z)$  is the logistic link function. The success probability  $\Pr(Acc_B = 1)$  is explained by the predictors  $t$  (transfer offered by Player  $A$ ) and  $c_B$  (costs of Player  $B$ ) and five treatment dummies  $D_1$  to  $D_5$ . The Maximum-Likelihood fit of (2) and some technical detail on the estimated model’s characteristics can be found in Appendix B.

Here, we start by using the fitted model to estimate the Average Marginal Effect ( $AME$ ) of each of the predictors, which quantifies their average isolated effect on success probabilities when controlling for all other variables.<sup>12</sup> The first two rows of Table 3 show that the average effect of  $B$ ’s costs on success probabilities is negative and significant ( $z$ -Test,  $p$ -value  $< 0.001$ ), whereas for the transfer  $t$  the  $AME$  is positive and significant. The next rows display the isolated effect of a treatment relative to a baseline both within and across the three settings. It is important to note that these comparisons differ conceptually from those carried out to quantify the “cooperation gaps” in Section 4.1. The cooperation gap is defined as the difference in the *acceptance rates* between two treatments. The acceptance rates themselves are not only influenced by whether there is a pre-commitment or not, but also by the transfers and

---

<sup>12</sup>See Long and Freese (2006) for details on the calculation of marginal effects of continuous and categorical predictors.

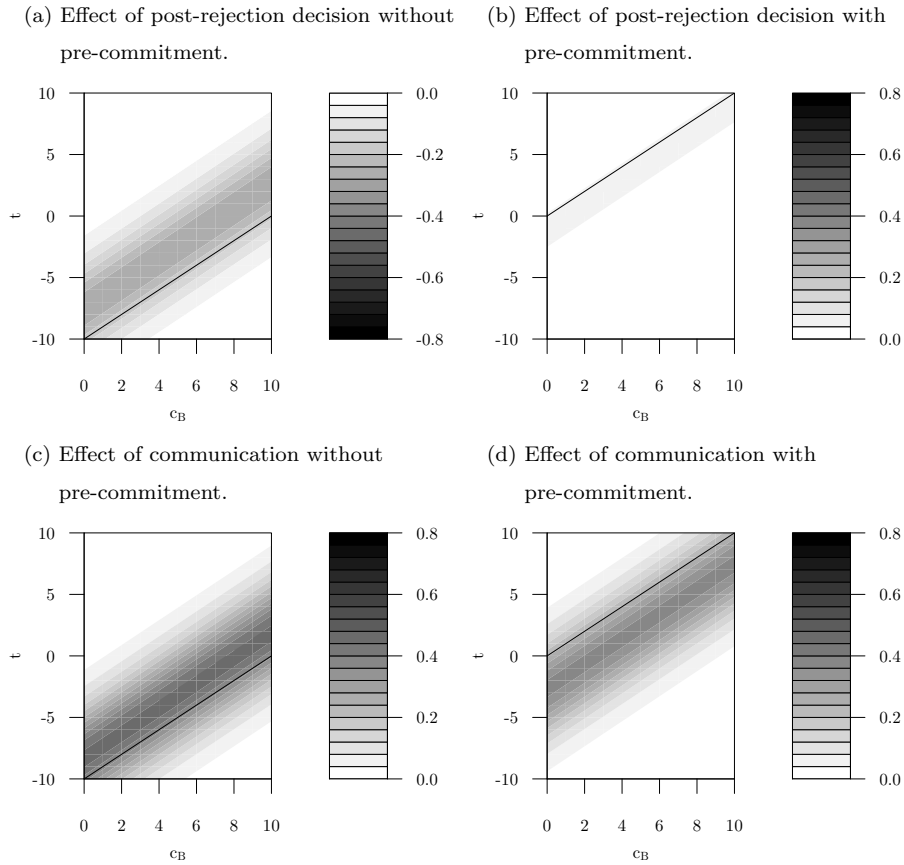
Table 3: Average Marginal Effects of continuous variables, within setting comparisons and between setting comparisons.

		<i>AME</i>	<i>Std.Err.</i>	<i>z</i>	$P >  z $	95% Conf. Int.	
						lower	upper
	$c_B$	-0.07	0.005	-14.27	< 0.001	-0.077	-0.058
	$t$	0.07	0.004	17.04	< 0.001	0.059	0.074
Baseline	vs.						
nPC_nD_nC	PC_nD_nC	-0.51	0.034	-15.05	< 0.001	-0.579	-0.445
nPC_D_nC	PC_D_nC	-0.38	0.038	-9.80	< 0.001	-0.450	-0.300
nPC_D_C	nPC_D_C	-0.35	0.038	-9.24	< 0.001	-0.427	-0.278
nPC_nD_nC	nPC_D_nC	-0.12	0.032	-3.62	< 0.001	-0.178	-0.053
PC_nD_nC	PC_D_nC	0.02	0.038	0.57	0.570	-0.053	0.096
nPC_D_nC	nPC_D_C	0.18	0.034	5.31	< 0.001	0.113	0.246
PC_D_nC	PC_D_C	0.20	0.041	4.95	< 0.001	0.122	0.282

costs that are present in each of the treatments being compared. Consequently, the evolution of  $AME$  values over the three settings (row 4 – 6 in Table 3) draw a different picture than the observed cooperation gaps. In our reference setting the  $AME$  of pre-commitment is very strong ( $AME = -0.51$ ) and becomes successively weaker under a post-rejection decision ( $AME = -0.38$ ) and communication ( $AME = -0.35$ ). In all three settings the isolated effect of a pre-commitment on success probabilities is negative and significantly different from zero.

The next four rows in Table 3 are across setting comparisons. We see that the introduction of the post-rejection decision has a significant negative effect without pre-commitment ( $AME = -0.12$ ), whereas under pre-commitment this effect disappears ( $AME = 0.02$ ,  $p$ -value = 0.57,  $z$ -Test). Finally, introducing communication has a positive and statistically significant effect without pre-commitment ( $AME = 0.18$ ) and with a pre-commitment ( $AME = 0.20$ ).

Figure 4: Treatment effects on estimated acceptance probabilities over all transfers  $t$  and costs  $c_B$ . Darker shadings represent stronger effects on the estimated acceptance probability. The black line represents theoretical  $t/c_B$ -combinations for which a payoff-maximizing  $B$  should be indifferent between accepting and rejecting.



Clearly, the averages of the marginal effects are useful, but the aggregation of multidimensional data into a single number always eliminates information. For example, *AMEs* hardly allow for drawing conclusions about non-payoff-maximizing behavior or to identify those combinations of transfer  $t$  and cost  $c_B$  for which a treatment effect is the strongest. For that reason we also plotted

the individual marginal treatment effects at every hypothetical combination of transfer  $t$  and cost  $c_B$ . Figure 4 visualizes this approach for each of the four across setting comparisons.

Figures 4a and 4b display the effect of a post-rejection decision with and without pre-commitment. Without pre-commitment we see a negative effect so introducing a post-rejection decision must have led some  $B$ -players to reject when they would have accepted otherwise. The solid black line represents all transfer/cost constellations for which  $B$  should be theoretically indifferent between accepting and rejecting. Above that line accepting results in higher payoffs, below it rejecting is more profitable. As the darkest shading (i.e. the strongest effect) is located strictly above that line a considerable number of  $B$ -player must have rejected even though it was not payoff-maximizing. This is in line with our previous conjecture that  $B$ -players began to develop expectations about exploiting the  $A$ -players in the prisoners' dilemma or simply wanted to punish  $A$  at cost to themselves.

In contrast to that, Figure 4b shows that, when  $A$  had to pre-commit, the effect of a post-rejection decision was miniscule but positive and is located strictly below the indifference line.<sup>13</sup> Put differently, if there was an effect at all under pre-commitment, then it was one of positive reciprocity, i.e. accepting non-payoff-maximizing offers.

Performing the same analysis for the effect of communication we observe a positive treatment effect from communication both without pre-commitment (Figure 4c) and with pre-commitment (Figure 4d). Communication must thus have made  $B$  players accept who would have rejected otherwise. Without pre-commitment those  $B$ -players who would have rejected without communication are those that should have accepted if they were payoff-maximizing because the darkest shading is located strictly above the indifference line. In other words, communication led to an increase in acceptance probabilities because it

---

<sup>13</sup>Note that the shading is so bright that it may become invisible in some printouts. We kept this color coding anyway to maintain comparability to the other three figures.

reduces negative reciprocity when no pre-commitment is involved. With pre-commitment those  $B$ -players who would have rejected without communication are those that should have rejected if they were payoff-maximizing because the darkest shading is located strictly below the indifference line. As communication made these players accept anyway, the increase in acceptance probabilities can be attributed to an increase in positive reciprocity when pre-commitment is involved.

## 5. Discussion

Our experimental results lend considerable support to the main predictions of the KT-model, even though subjects in our experiment were influenced by behavioral motives that are not taken into account in the original model. The experiment revealed behavior consistent with a tendency towards cooperation, positive and negative reciprocity, and inequality aversion. All these traits can work against Result 1 of the model. Cooperative behavior and positive reciprocity can lead to more agreements under pre-commitment and negative reciprocity to less agreement without pre-commitment. So if subjects had been led more by reciprocity or social preferences than by payoff-maximizing behavior the theoretical results could well have failed to hold. Nevertheless, in all our comparisons we observed that a pre-commitment reduced the likelihood of an agreement being reached.

Communication between the players also had an interesting effect. As we have come to expect from past experiments communication led to subjects having a higher willingness to deviate from payoff-maximization and behave cooperatively. Given this, we expected the treatment with communication to strengthen the effects such as cooperative behavior and reciprocity that can work against the model's results. However, while communication did have the expected effect on the willingness to cooperate it also led to a sharp fall in negative reciprocity and on the attempts of the  $B$ -players to exploit the  $A$ -players in the treatment without pre-commitment. Combined this resulted in the number



of agreement being reached being the highest of all treatments in the no pre-commitment treatment with communication. This in turn led to the difference between pre-commitment and no pre-commitment increasing with communication and the difference once again being statistically significant.

All in all, our experimental results provide important backing to the main conclusion of the KT-model regarding the potentially counterproductive effects of pre-commitments in climate negotiations. We have shown that even if negotiating parties are driven by behavioral motives outside the model the results of the model can still be expected hold. Moreover, we have shown that if the environment is extended to include the realistic feature of communication between parties the main conclusion still holds strong.

A possible critique of the KT-model – and thus our experimental investigation – is that the pre-commitment of the *A*-players is not voluntary. It could be argued that in the case of voluntary pre-commitments positive reciprocity would be more pronounced. There are good reasons for not testing this by running experiments with voluntary pre-commitments using the KT-model. First, given the specific bargaining situation, we cannot expect to observe any voluntary pre-commitment. If *A*-players have the choice between committing and not committing, they will certainly opt for not committing, even if they harbor some kind of other regarding preferences. The reason is that to pre-commit actually increases the inequality of final payoffs.

Therefore, an inequality averse player would also choose to not pre-commit and would offer a low price ( $t = 0$ ) which ensures that the *B*-player will agree to cooperate and which leads to equal payoffs for both players. Consequently, there is no motivation discussed in the literature on “other regarding preferences” which would make pre-commitment a rational choice.

This goes in line with the fact that we rarely observe truly altruistic pre-commitments in reality. In the context of climate damage abatement, for example, the pre-commitment is usually sold as having the advantage of allowing countries that pre-commit to gain a competitive advantage in the development of cleaner technologies and not out of some altruistic consideration.

Furthermore, previous experiments on the effect of pre-commitments in which these were voluntary have shown little evidence for reciprocity which go beyond that which we observed in our experiment. In the literature on leadership in climate negotiations we have seen only moderate levels of reciprocal behavior towards pre-commitments (Güth et al., 2007; Levati et al., 2007; Gächter et al., 2012; Sturm and Weimann, 2008). In a sequential bargaining experiment Brosig et al. (2004) find that voluntary pre-commitments are greatly taken advantage of and in a current paper Heinrich and Weimann (2014) find that in dictator games in which recipients could choose between different modified dictator games there was no reciprocity shown by dictators. Thus, there is no current experimental evidence that suggest a voluntary pre-commitment would lead to more reciprocity than that which we observed in our experiment.

### **Acknowledgement**

We are grateful to Annette Kirstein, Michael Kvasnicka, Hendrik Thiel, the participants of the 2013 meeting of the Social Science Commission/Verein für Socialpolitik, the participants of the 2013 Economic Science Association World Meeting, and two anonymous reviewers for their valuable feedback.

### **References**

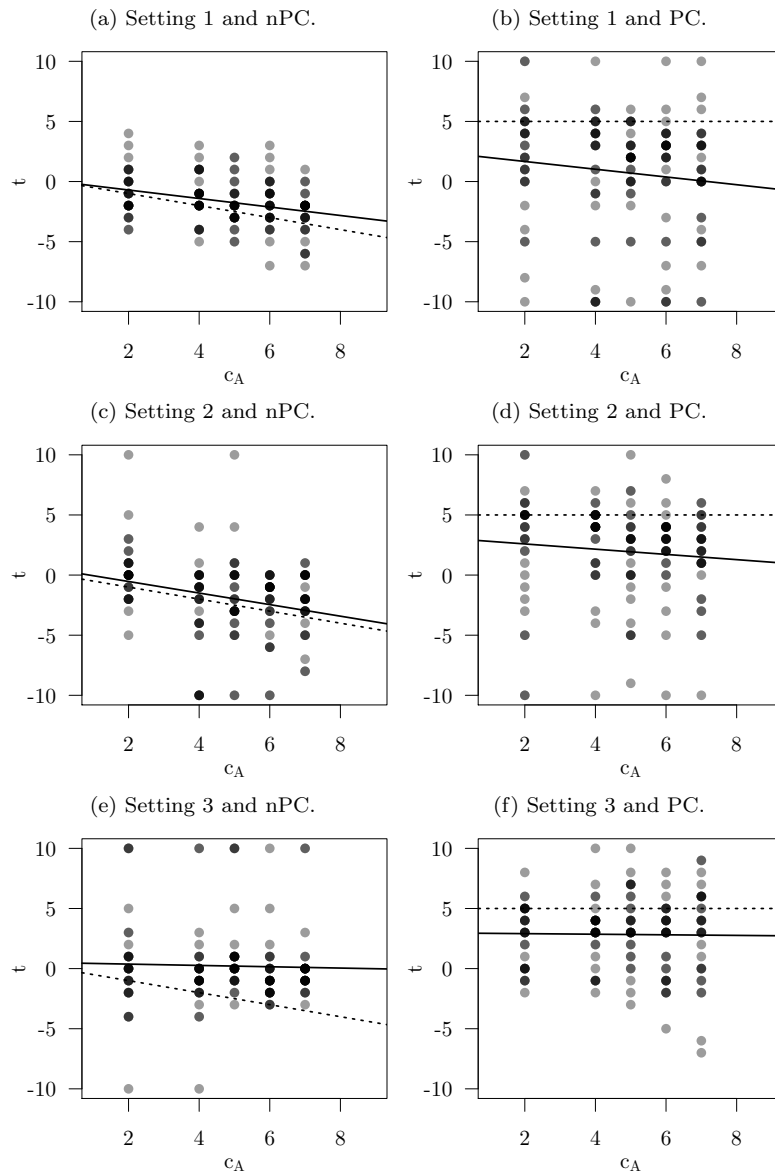
- Barrett, S., 1998. On the theory and diplomacy of environmental treaty-making. *Environmental and Resource Economics* 11, 317–333.
- Bolton, G.E., Ockenfels, A., 2000. Erc: A theory of equity, reciprocity, and competition. *American Economic Review* 90, 166–193.
- Brosig, J., Weimann, J., Ockenfels, A., 2003. The effect of communication media on cooperation. *German Economic Review* 4, 217–241.
- Brosig, J., Weimann, J., Yang, C.L., 2004. Communication, reputation, and punishment in sequential bargaining experiments. *Journal of Institutional and Theoretical Economics JITE* 160, 576–606.

- Carraro, C., Siniscalco, D., 1993. Strategies for the international protection of the environment. *Journal of Public Economics* 52, 309–328.
- Falk, A., Fischbacher, U., 2006. A theory of reciprocity. *Games and Economic Behavior* 54, 293–315.
- Fehr, E., Gächter, S., 2000. Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14, 159–181.
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* 114, 817–868.
- Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10, 171–178.
- Gächter, S., Nosenzo, D., Renner, E., Sefton, M., 2012. Who makes a good leader? social preferences and leading-by-example. *Economic Inquiry* 50, 867–879.
- Gresik, T., Satterthwaite, M., 1989. The rate at which a simple market converges to efficiency as the number of traders increases: An asymptotic result for optimal trade mechanisms. *Journal of Economic Theory* 48, 304–332.
- Güth, W., Levati, M.V., Sutter, M., Van Der Heijden, E., 2007. Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics* 91, 1023–1042.
- Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3, 367–388.
- Heinrich, T., Weimann, J., 2014. A note on reciprocity and modified dictator games. *Economics Letters* 121, 202–205.
- Kamecke, U., 1997. Rotations: Matching schemes that efficiently preserve the best reply structure of a one shot game. *International Journal of Game Theory* 26, 409–417.

- Konrad, K., Thum, M., 2014. Climate policy negotiations with incomplete information. *Economica Wiley Early view* (Online version), doi: 10.1111/ecca.12065.
- Lange, A., Löschel, A., Vogt, C., Ziegler, A., 2010. On the self-interested use of equity in international climate negotiations. *European Economic Review* 54, 359–375.
- Lange, A., Vogt, C., Ziegler, A., 2007. On the importance of equity in international climate policy: An empirical analysis. *Energy Economics* 29, 545–562.
- Levati, M.V., Sutter, M., Van der Heijden, E., 2007. Leading by example in a public goods experiment with heterogeneity and incomplete information. *Journal of Conflict Resolution* 51, 793–818.
- Long, J.S., Freese, J., 2006. Regression models for categorical dependent variables using Stata. 2 ed., Stata Press.
- Meyerson, R., Satterthwaite, M., 1983. Efficient mechanisms for bilateral trading. *Journal of Economic Theory* 29, 265–281.
- Nowak, M.A., Roch, S., 2006. Upstream reciprocity and the evolution of gratitude. *Proceedings of the Royal Society* 274, 605–609.
- Nowak, M.A., Siegmund, K., 2005. Evolution of indirect reciprocity. *Nature* 437, 1291–1298.
- Rob, R., 1989. Pollution claim settlements under private information. *Journal of Economic Theory* 47, 307–333.
- Sturm, B., Weimann, J., 2008. Unilateral emissions abatement: An experiment, in: *Environmental economics, experimental methods*, Routledge. pp. 157–183.
- Valley, K., Moag, J., Bazerman, M.H., 1998. A matter of trust: Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior and Organization* 34, 211–238.

## Appendix A. Estimated relationships between transfer and A's costs

Figure A.5: Estimation of linear relationships between the A's transfers and costs. The darker the shading of dots the more observations there are. Dashed lines represent relationships predicted by the KT-Model, solid lines are the fitted ones.



In what follows we present the regression results of the linear model  $t = \alpha + \beta c_A$  for all six treatments. Figure A.5 shows scatterplots of our experimental data. Each the three rows in that figure represent one of the Settings 1 – 3. In the left column we have treatments without pre-commitment (**nPC**) and the right column shows treatments with pre-commitment (**PC**). The estimated coefficients and  $p$ -values of comparative hypotheses in the **nPC**-condition are given in Table A.4.

Table A.4: Parameter estimates and  $p$ -values of the linear relationship between  $t$  and  $c_A$  for all three settings without pre-commitment.

<b>nPC</b>	$\hat{\alpha}$	$\hat{\beta}$	$p$ -values		
			$H_0 : \alpha = 0$	$H_0 : \beta = -\frac{1}{2}$	$H_0 : \alpha = 0 \text{ and } \beta = -\frac{1}{2}$
Setting 1	0.002	-0.353	0.99	0.11	< .001
Setting 2	0.427	-0.481	0.58	0.90	0.14
Setting 3	0.484	-0.055	0.55	0.01	< .001

## Appendix B. Logistic regression results

Table B.5: Logistic regression with `nPC_nD_nC` as baseline. *Modell characteristics:* (1) No multicollinearity: All variance inflation factors are  $\ll 10$ , (2) No perfect separation: All variables converged after `nsteps=4` using `separation.detection()` in R Package `brglm`, (3) Null deviance: 1193.3 on 899 degrees of freedom, Residual deviance: 671.5 on 892 degrees of freedom, Likelihood-Ratio test for overall fit:  $\chi^2 = 521.8$ ,  $p < 0.001$  (4) Predictive accuracy: In 84.5% of all cases the observed 1 (0) was predicted at  $p \geq 0.5$  ( $p < 0.5$ ).

	$\hat{\beta}$	$\hat{sd}$	Wald $\chi^2$	$p$	Odds Ratio	95% Conf. Int.	
						lower	upper
(Intercept)	4.81	0.40	146.20	< 0.001	–	–	–
Transfer <code>t</code>	0.56	0.05	135.38	< 0.001	1.76	1.60	1.94
Cost <code>cB</code>	–0.57	0.05	110.69	< 0.001	0.57	0.51	0.63
<code>PC_nD_nC</code>	–4.26	0.42	101.02	< 0.001	0.01	0.01	0.03
<code>nPC_D_nC</code>	–1.13	0.32	12.81	< 0.001	0.32	0.17	0.60
<code>PC_D_nC</code>	–4.07	0.42	92.60	< 0.001	0.02	0.01	0.04
<code>nPC_D_C</code>	0.84	0.39	4.54	0.033	2.32	1.09	5.17
<code>PC_D_C</code>	–2.49	0.39	41.95	< 0.001	0.08	0.04	0.17

## Appendix C. Analysis of communication

A detailed look at how the communication was used in the two treatments can shed some light on the widening of the cooperation gap. Table C.6 shows how the chats were used in the two treatments with communication. As can be seen cost information was revealed in more cases in the `nPC_D_C`-treatment (71%) than in the `PC_D_C`-treatment (42%) with the share of truthful revelation being roughly the same between the two (60% vs. 69%). After there was a

communication about costs there was a strong tendency to reach an agreement. Furthermore, the tendency was stronger in the `nPC_D_C`-treatment than in the `PC_D_C`-treatment (94% vs. 72%). Thus, the fact that both costs were revealed more often and that the tendency to reach agreement after a cost revelation was stronger in the `nPC_D_C`-treatment is consistent with the widening of the cooperation gap.

Table C.6: Communication in Setting III

Cases in which...	nPC	PC
... <i>A</i> and <i>B</i> revealed own cost in chat	214 (71.3%)	127 (42.3%)
... revealed cost was true value	129 (60.3%)	87 (68.5%)
... <i>A</i> and <i>B</i> agreed after costs were revealed	202 (94.3%)	91 (71.7%)
... <i>A</i> and <i>B</i> agreed in chat	194 (64.7%)	160 (53.3%)
... <i>A</i> and <i>B</i> agreed in game	189 (97.4%)	144 (90.0%)

In the `nPC_D_C`-treatment 65% of the the subjects made a firm commitment to come to an agreement in the chat while in the `nPC_D_C`-treatment it was 53%. Again, there was a strong tendency of offers to be accepted after a clear agreement was reached in the chat (97% and 90%). This difference between agreements reached in the chats is again consistent with a widening of the cooperation gap.

#### Appendix D. Instructions

The following text represents the instructions in the `nPC_D_nC`-treatment Treatment (translated from German):



## **i** Rules and instructions, please read carefully!

1. **Before the experiment** switch-off your mobile phone. Read carefully the instructions below and inform the experimenter by a show of hands if you have questions.
2. **During the experiment** you are neither allowed to talk nor to leave your place. We let you know when the experiment is finished. If your screen does not respond to your entry immediately, an other player has not decided yet. Please be patient.
3. **After the experiment** you get your payoff. Please remain seated until you are called.

**General procedure:** Before the actual experiments you will play 3 practice rounds against the computer. These practice rounds are not payoff relevant. The purpose of these rounds is to familiarize yourself with the experimental environment. The computers behavior will not be arbitrary but fully payoff-maximizing. In order that you are fully able to comprehend the computers actions you will be given the computers private cost information in the first two practice rounds. This information will be unknown to you both in the third practice round as well as in the actual game.

In the following 5 rounds you will play the actual experiment in which you will play each other participant once. At the start you will receive an endowment which can grow or shrink depending on your actions in each of the rounds. Your final payoff after the 5 rounds will certainly be positive. There are a total of 10 participants, 5 of which will be given the role of an *A* player and 5 the role of a *B* player. In each round the first decision will always be made by *A* and the second by *B*. The *B* player will thus be able to observe the action taken by the *A* player before making a decision. Designated roles will stay the same throughout the experiment.

**Content of the experiment:** In each of the rounds you and the other player can contribute some effort  $e$  towards the provision of a common good. The size of  $e$  is measured in lab dollars with 5 lab dollars being worth 1 euro. The choice  $e = 10$  means that you will contribute the effort, and  $e = 0$  means you will not contribute the effort. Other values of  $e$  are not possible. You and the other player will receive the sum of both your contributions. That is, 0 (if neither of you contributed), 10 (if one of you contributed), 20 (if you both contributed). This payoff will be reduced by your cost of contributing. The cost will be zero if you do not contribute ( $e = 0$ ) and  $10 + c$  if you do contribute ( $e = 10$ ). The whole number  $c$  is randomly selected and lies between 1 and 9, with each of the nine values being equally probable. Before each round each player will find out their own cost but not the cost of the other participant.

The *A* player can offer the *B* player a transfer payment which can take any whole number between  $-10$  and  $+10$ . Positive values mean there is a payment from *A* to *B* and negative values mean there is a payment from *B* to *A*. Player *B* can either accept or reject the transfer offer. If the *B* player accepts the transfer payment will be made and both players are required to contribute the effort  $e = 10$  (you will then no longer be able to decide over this freely as  $e = 10$  will be set automatically). If *B* rejects the offer then both players can choose freely whether to contribute or not, i.e., you can either choose  $e = 10$  or  $e = 0$ .

**Experimental timeline::**

- **Stage 1:** Player *A* receives the cost information  $10 + c_A$  and offers *B* a transfer  $-10 \leq t \leq 10$
- **Stage 2:** Player *B* receives the cost information  $10 + c_B$  and decides whether to accept the offer or not.
- **Acceptance, Stage 3a:**  $e = 10$  is set for both players.
- **Rejection, Stage 3b:** Both player decide between  $e = 10$  and  $e = 0$ , without knowing the decision of the other player.

After Stage 3 the current round ends and you are informed about your current payoff. Payoffs from previous rounds are logged and made available to you in each round.

**Payoffs:** Lab-dollars  $\pi$  in a round are determined as follows (**red** for **Player A**, **blue** for **Player B**):

- If *B* accepts (Stage 3a):

	gross payoff	-	costs	±	transfer
$\pi_A =$	20	-	$(10 + c_A)$	-	$t$
$\pi_B =$	20	-	$(10 + c_B)$	+	$t$

- If *B* rejects (Stage 3b):

effort		B		
		$e_B = 0$	$e_B = 10$	
A	$e_A = 0$	0	10	$-c_B$
	$e_A = 10$	$-c_A$	$10 - c_A$	$10 - c_B$

## Appendix E. Screenshots

Figure E.6: Screen of Player A in Setting 2 and nPC.

**Your cost:**  
15

**Your income at different transfer amounts**

Transfer:	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
Your income:	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	-1	-2	-3	-4	-5

**Player B's income at different transfer amounts**

Transfer:	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
B's income:	-Cb	1-Cb	2-Cb	3-Cb	4-Cb	5-Cb	6-Cb	7-Cb	8-Cb	9-Cb	10-Cb	11-Cb	12-Cb	13-Cb	14-Cb	15-Cb	16-Cb	17-Cb	18-Cb	19-Cb	20-Cb

**Possible income if offer is rejected**

	You both choose 0	You choose 10 and player E 0	Player B chooses 10 and you 0	You both choose 10
Your income	0	-5	10	5
B's income	0	10	-Cb	10 - Cb

How much do you wish to offer?

Figure E.7: Screen of Player *B* in Setting 2 and nPC.

**Your cost:**  
17

**Player A's offer:**  
4

**Your income if you accept**  
7

**Player A's income at different cost if you accept the offer**

A's possible costs:	1	2	3	4	5	6	7	8	9
A's income:	5	4	3	2	1	0	-1	-2	-3

**Possible income if offer is rejected**

	<b>You both choose 0</b>	<b>You choose 10 and player A 0</b>	<b>Player A chooses 10 and you 0</b>	<b>You both choose 10</b>
<b>Your income</b>	0	-7	10	3
<b>A's income</b>	0	10	-Ca	10 - Ca

**Do you accept the offer?**

Accept the offer  
 Reject the offer