Reciprocity Reciprocity in Climate Coalition Formationin Climate Coalition Formation

Lin, Yu-Hsuan

Department of Economics, the Catholic University of Korea

2018

Reciprocity in Climate Coalition Formation

Yu-Hsuan Lin[1]

Working Paper

Version: March 2018

[**Abstract**]

This study investigates the impact of reciprocal altruistic attitudes on individual willingness to participate in a climate coalition with experimental evidences. The theoretical result suggested that the scope of the coalition's formation could be enlarged by the participation of altruists. However, we found that a kind participant in the altruism test could behave unkindly to others in the public good game. Considering attitudes against reciprocal altruism, when participants thought they were being treated badly, they were more likely to join a coalition because of the threat of punishment. In contrast, when participants were noncritical to a coalition, such altruistic attitudes were insignificant to their decisions. This result implies that decisions in international conventions are not self-enforced. Overall, this study reveals that self-interest remains the key factor influencing individual participation in climate coalitions. Coalition formation can also be affected by reciprocal altruistic preferences.

Keywords: social preference, experimental design, reciprocity, altruism, international environmental agreements

[1] Department of Economics, the Catholic University of Korea. 43 Jibong-ro, Wonmi-gu, Bucheon-si, Gyeonggi-do, 14662, Korea. E-mail: yuhsuan.lin@catholic.ac.kr

This study is supported by the Catholic University of Korea Research Fund.

1 Introduction

Since Barrett (1994), a large amount of literature (such as (Bahn, Breton, Sbragia, & Zaccour, 2009; Barrett, 2001; Bratberg, Tjøtta, & Øines, 2005; Breton, Sbragia, & Zaccour, 2010; Finus & Rübbelke, 2013)) has focused on the formation of international environmental agreements (IEAs). The existing theoretical literature suggests that a large-scale IEA cannot usually exist without a policy mechanism. However, experimental studies (Burger & Kolstad, 2010; Kosfeld, Okada, & Riedl, 2009; Willinger & Ziegelmeyer, 2001) suggest that high levels of cooperation do exist in the absence of policy interventions. Such studies have claimed that people are far less likely to offer a free ride and more likely to cooperate than the Nash prediction suggests. Therefore, social (or other-regarding) preferences have been proposed by recent studies (such as Charness and Rabin (2002) and Dannenberg, Löschel, Paolacci, Reif, and Tavoni (2015)) to address this knowledge gap.

While social preferences have received some attention from economists, unidirectional social preferences have been discussed. For example, simple altruism proposed that people may care not only about their own well-being but also about the well-being of others. Such concern for others is unidirectional and does not ask for anything in return. Yet, psychological evidence indicates that most altruistic behavior is more complex (Rabin, 1993). People make decisions based on how they are treated by others: when they meet altruistic people, they are generous; when they meet unkind people, they are mean.

The motivation underlying reciprocal behavior has been discussed. Such studies can be categorized into three groups depending on motivations: reciprocal fairness (e.g., Bolton and Ockenfels (2000); Fehr and Schmidt (1999)), reciprocal altruism (Levine, 1998) (Dufwenberg & Kirchsteiger, 2004), and the quest for efficiency gains through (Brandts & Schram, 2001). There are also overviews of these (Seinen & Schram, 2006). In this paper, we intend to contribute to discussions of the motivations underlying reciprocal altruism.

In a previous study, Lin (2017) considered the impact of a unidirectional altruistic preference for the formation of a climate coalition. The study presumed that individual altruistic attitude is consistent in both dictator and public-good games. However, the design neglected bilateral social preferences. The experimental result showed that a player could be kind to others in a unidirectional dictator game but unkind in an interactive public-good game. In order to understand the cooperative behavior of individuals even when it is not in their interest to cooperate, a further investigation is required.

This study asks the following research questions: (1) is individual social preference unidirectional or mutual? (2) Is coalition formation affected by individual social preferences? The first question asks about the magnitudes of unidirectional and mutual impacts at an individual level. The second question asks about individual impacts at a global level.

This study examines individual decisions to participate in an IEA that is based on a model of reciprocity. We investigate the role of reciprocal preferences in a climate negotiation. The literature has shown that a grand coalition or a majority coalition may be stabilized by sufficiently strong and widespread reciprocity preferences. However, it is questionable whether this remains realistic in practice. This study tests that hypothesis using the existing experimental evidence.

This study employs Rabin (1993)'s framework of reciprocity, which considers a player's own payoff, the player's perception of others' payoffs, and others' perception of the player's payoff. When a player concerns about others' payoffs, she/he has more incentive to participate in a coalition. The coalition is therefore enlarged. However, given reciprocity preferences, a coalition may become smaller or larger than the prediction with self-interest preference due to the interactive perceptions of players' payoffs. This study will support that result using both theoretical predictions and experimental evidence, and the findings of this study will contribute to practical climate negotiations.

This study contributes to the understanding of climate coalition formation. Particularly, we examine how individual reciprocity preferences affect a coalition. We find that individuals have reciprocal altruism toward others. In other words, their decisions depend on how they are treated by others.

This paper is structured as follows: section 2 describes a climate negotiation using the model of reciprocity. Section 3 tests the model using experimental evidence. Section 4 compares the theoretical and experimental results. The final section concludes.

2 Model of Reciprocity

Consider a two-player reciprocity model, following Rabin (1993), with strategy sets S_1 and S_2 for players 1 and 2. The material payoff of a player π_i depends on the strategies chosen by both players ($S_1 \times S_2$). We use the following notation:

$a_1 \in S_1$ and $a_2 \in S_2$ represent the strategies chosen by the two players; $b_1 \in S_1$ and $b_2 \in S_2$ represent player 1′s belief about which strategy player 2 is choosing and player 2′s belief about which strategy player 1 is choosing; $c_1 \in S_1$ and $c \in S_2$ represent player 1′s belief about what player 2 believes player 1′s strategy is, and player 2′s belief about what player 1 believes player 2′s strategy is. In practice, we assume that b_1 and b_2 are player 2′s and player 1′s decisions in the past, respectively; and c_1 and c_2 are player 1′s and player 2′s decisions in the past, respectively.

We assume that each player′s subjective expected utility depends on three factors: (i) player i′s own interest, (ii) the other player j′s interest and (iii) the player i′s interest influenced by the other player j′s strategy. The first factor is a linear function of player i′s own payoff. The second and third factors incorporate direct kindness and reciprocal kindness. Direct kindness indicates player i′s kindness to player j, while reciprocal kindness indicates how player i experiences player j′s kindness. Both kindness values are defined as follows:

Definition 1: Player i′s kindness to player j is given by

$$u_i^a(b_j, a_i) \equiv \frac{\pi_j(b_j, a_i) - \pi_j^e(b_j)}{\pi_j^h(b_j) - \pi_j^{min}(b_j)} \tag{1}$$

If $\pi_j^h(b_j) - \pi_j^{min}(b_j) = 0$, then $u_i^a(b_j, a_i) = 0$.

where a_i represents the strategies chosen by player i, and b_j represents player i′s belief about which strategy player j is choosing. Player i′s kindness depends on four payoffs: player j′s payoff π_j ; $\pi_j^h(b_j)$ is player j′s lowest payoff among points that are Pareto-efficient in the set of all players′ possible payoffs that player i believes player j is choosing; $\pi_j^e(b_j)$ is the equitable payoff and is defined as the average of the highest and lowest payoffs, $[\pi_j^h(b_j) + \pi_j^l(b_j)]/2$; and $\pi_j^{min}(b_j)$ is the worst possible payoff for player j in the set of all possible payoffs.

Definition 2: Player $i's$ belief about how kind player j is being to her is given by

$$u_i^r(b_j, c_i) \equiv \frac{\pi_i(c_i, b_j) - \pi_i^e(c_i)}{\pi_i^h(c_i) - \pi_i^{min}(c_i)} \tag{2}$$

If $\pi_i^h(c_i) - \pi_i^{min}(c_i) = 0$, then $u_i^r(b_j, c_i) = 0$.

Both straight kindness and reciprocal kindness are in the range of -1 and $1/2$.

Therefore, we assume that player i′s subjective expected utility is:

$$u_i = u_i^s(\pi_i\ ) + u_i^r(\pi_i\ , \pi_j\ )[1 + u_i^a(\pi_i\ , \pi_j\ )] \tag{3}$$

where the first term u_i^s is pure self-interest, the second term contains two parts: first part u_i^r is reciprocal altruism and the latter one u_i^a is straight altruism. Reciprocal kindness plays an important role in utility. If the player feels being treated badly, her overall utility would be lower than her monetary payoff. On the other hand, straight kindness plays the role of strength of feeling. If a player is straightly unkind, she cares less about others' decisions. But if a player is straight generous, her utility depends on how she is treated: if she is treated kindly, her utility is higher; if she is treated badly, her utility is lower than the monetary payoff.

Turning to the membership game with n players, the players were simultaneously asked to decide whether to participate in a coalition. Following D'As premont, Jacquemin, Gabszewicz, and Weymark (1983), when individuals were characterized by reciprocal preference, a stable coalition existed when the internal and external constraints were satisfied as follows:

$$u^n\,(n^*) > u^{ns}(n^* - 1) \tag{4-1}$$

$$u^{ns}(n^*) > u^n\,(n^* + 1) \tag{4-2}$$

where the superior letters n and ns denote a player's membership status: n means signatory and ns means nonsignatory, the number in the bracket is the number of signatories. The internal constraint (4-1) denotes that a player is happier to be a part of the n^*-th coalition. If that constraint is satisfied, then all signatories have no incentive to leave the coalition. The external constraint (4-2) indicates that a non-signatory does not wish to participate in a coalition as the $(n^* + 1)$-th member. If this constraint is satisfied, then all of the non-signatories do not want to participate. When both constraints are satisfied, a coalition is stable with n^* signatories

3 Experimental results

We employ the experimental results from Lin (2017). The experiment was conducted at the center for Experimental Economics laboratory at the University of York (UK) in 2013 and programmed with z-Tree ((Fischbacher, 2007)). Fifty students with multi-cultural and multi-disciplinary backgrounds were invited to mimic a climate negotiation. However, any content related to environmental issues was excluded from the instructions to avoid biases due to subjects' attitudes towards the environment. Each subject had to complete two parts in this experiment: the first part examined individual altruistic attitudes, and the second part measured interactive social preferences in a public good game. In order to purify the research goal, there was no environmentally related content in the instructions. The experiment took place as follows.

Prior to the experiment, a questionnaire was circulated to gather demographic information, including the subject's degree discipline, age, ethnicity, political orientation and level of belief in a religion. The demographic results show the diversity of the subjects' backgrounds. The distribution of their levels of belief in religion was on a scale ranging from not religious to extremely religious. The results showed that 40% considered themselves atheist. Meanwhile, 6, 8, and 9 subjects considered themselves as atheists with mild, medium, and strong beliefs, respectively; and 7 considered themselves pure religionists. The average level of 2.5 implies that the strength of the subjects' religious beliefs ranged from mild to medium. The other question aimed to capture their political preferences, ranging from left-wing to right-wing. The distribution showed that 7 subjects identified themselves as left-wing; 10 as center-left; 25 as neutral; 7 as center-right and 1 as right-wing. To ensure data quality, the subjects had to understand the rules of the game as much as possible. Thus, the experimenter introduced the rules and gave the subjects time to read through the instructions thoroughly and to complete the controlled questions. At the end of each part of the experiment, four control questions were asked to test the subjects' understanding of the payoff tables. A new part would only start if all subjects had answered all control questions correctly.

In the altruism test, subjects were anonymously and randomly paired with each other to make 20 '*keep*' or '*give*' decisions. In each round, each subject was given 1 token, and they decided whether to give it to their partners. On the other hand, they did not know their partner's decision until the end of the session. In the 20 rounds, there were different monetary values for keeping and giving the token.

Next, a public good provision game was conducted. Subjects were randomly assigned to groups of 5 persons for the whole session, which was conducted anonymously. Each subject had a payoff table for all 26 possible coalition combinations. Payoffs for each player were not identical. The players' payoffs, in the range of £0 to £24, depended on their decisions and the combination of players in the coalition. They were asked to make a decision to join or not join a coalition for 4 different treatments in 60 rounds. In contrast to the altruism test, at the end of each round, subjects were informed about their own payoffs, other group members' decisions, and the coalition formation.

In particular, the experiment developed treatments with a self-interested dominant strategy equilibrium condition. Each player had a clear dominant strategy for whether to participate in a coalition. Players were divided into two groups: *critical* players, who were essential to an effective coalition, and *noncritical* players, who were able to free-ride the public good benefits. The condition implies that any critical country could not be replaced by the joint of noncritical players. In other words, critical players would participate in a coalition because they were necessary members, and noncritical players would not participate because of the free-riding advantage. The condition ensures that the formation was the only stable effective coalition.

While we acknowledge this is indeed a strong condition, in order to identify the individual incentives to participate in the coalition, this condition offered the primary strength of investigating individual incentives for participating in IEAs. If there was more than one stable coalition, the individual decisions were difficult to predict. However, when there was a dominant strategy equilibrium, it provided a suitable environment in which to observe individual decisions when every player had an optimal strategy to choose.

Tables 1A and 1B show each player's marginal benefits in eight treatments. Five subjects played four treatments in 60 rounds. Their payoffs depended on the individual marginal benefit of the total contribution: a signatory's payoff is the marginal benefit times the summation of all signatories' marginal benefits, minus the participation cost; a non-signatory's payoff is the marginal benefit times the summation of all signatories' marginal benefits. The treatments were designed for stable coalitions of 2 to 4 critical players. As explained earlier, based on the assumption of self-interest, the dominant strategy equilibrium design could help to identify individual decisions. Critical players were essential for an effective coalition, while noncritical players had the incentive to free ride.

When the subjects had strong inequality-averse attitudes, then the critical players might have had the incentive to break the coalition internally. On the other hand, noncritical players might have given up the free-riding benefit by participating in a coalition. In this study, we assigned each subject a particular payoff table, which contained all of the possible payoffs with the corresponding coalition combinations[2]. The payoff depended on the given parameters and the coalition formation. For any ineffective coalition, all of the subjects in the group gained nothing in return. The

[2] A possible coalition combination requires at least 2 players. Therefore, there were $(2^5 - 5 - 1) = 26$ possible coalition combinations.

possible payoffs for the subjects ranged from £0 up to £24.

4 Analyses with experimental evidence

Table 2 reports the number of tokens subjects decided to give when the values of keeping and giving the tokens varied. As mentioned earlier, different monetary values were attributed to keeping and giving the tokens (z_1 and z_2 respectively).

The ratio of keeping-to-giving values (z_1/z_2) was designed from 1 to 0.05 in 20 rounds. In this altruism test, it is perhaps unsurprising that no subject decided to give his/her tokens in the first round. However, when the ratio of keeping-to-giving values became smaller, more and more subjects would give their tokens away. In the final round, nearly 60% of subjects gave up the token for £0.5 to allow a stranger to earn £10.

In general, an increasing trend is noticed, such that subjects become altruistic when the token was more valuable to receivers than to givers. This interesting point shows that the value to the giver is an important factor in a subject's decision-making. When the value of the token to a giver was small (e.g., rounds 8, 14, and 17), subjects were more likely to behave altruistically by giving it up.

In the simple altruism test, subjects did not know how they were treated by their partners. Therefore, only pure altruism – but no reciprocal altruism – was calculated. Because the decisions of both keeping and giving are Pareto-efficient solutions, the lowest and worst payoffs for other players were the same (i.e., not giving to the partner). A subject's direct altruism level is either -0.5 (keep) or 0.5 (give). The decisions in 20 rounds indicate the subject's direct altruism level. The mean value of all subjects' altruism levels is (-0.21), which means that subjects were unkind to others in general.

Table 3 shows the OLS estimation of pure altruism attitudes. The dependent variable is the individual's average direct altruism level. Independent variables are the factors selected from the questionnaire, including subjects' ages, political attitudes, and religious attitudes. The results show that only religious attitude is significant for the subject's altruism, at a 10% level. In other words, subjects with stronger religious beliefs behaved more altruistically toward others.

Turning now to coalition formation in the membership game, effective coalitions were formed in 387 out of 600 rounds, and the formation was usually larger than the self-interested equilibrium size. The actual coalition formation matched the self-interested equilibrium in only 112 rounds. The coalitions were usually neither stable nor convergent to a particular coalition. With the same treatments, the coalition formation varied in different groups. For example, group 6 and group 8 both took treatments 5 to 8. Group 6 formed profitable coalitions in 47 rounds, but group 8 achieved effective coalitions in only 12 rounds.

We now examine the factors that might affect individual decisions; Maximum Likelihood Estimation (MLE) of binary probit regressions was employed in Table 4. The variables include (v1) the decision made in the previous round, (v2) the year the subjects were born, (v3) political attitudes from left to right, (v4) religious attitudes from atheist to religionist, (v5) the dummy variable of being critical players, (v6) the marginal benefit of total contribution, (v7) the group contribution in the previous round and (v8) individual attitudes in the pure altruism test.

Individuals' direct altruism and reciprocal altruism are measured as follows. Considering definitions 1 and 2, we employ the decisions made in the prior round to indicate a player's kindness toward other players. Hence, in equations (1) and (2), b_1 and b_2 are player 2's and player 1's decisions in the past, respectively; and c_1 and c_2 are player 1 and player 2's decisions in the past, respectively.

The highest and lowest Pareto-efficient payoffs are the highest and lowest payoffs in a possible effective coalition. For a critical player, the highest payoff is a grand coalition solution and the lowest Pareto-efficient payoff is the self-interested Nash solution. For a noncritical player, the highest payoff is a solution in which the player is the only nonsignatory, and the lowest Pareto-efficient payoff occurs when the player joins with the critical players only. In addition, the worst payoff occurs when no coalition is formed (everyone gets nothing). Because a player faces 4 other players in a group, her altruism is the average of her altruism toward other players. Similarly, by using the players' historical decisions, we can determine a player's reciprocal altruism. In other words, a player's reciprocal altruism is her sense of how kind other players were being to her.

A subject's altruism attitudes were the average of her attitudes to other four players in the group. We obtain 2,800 samples (due to the exclusion of the first observation in every treatment) to show players' altruism attitudes. A correlation coefficient was computed to assess the relationship between an individual's pure and reciprocal altruism. There was a strong and positive correlation (0.84) between pure and reciprocal altruism. The mean values of direct and reciprocal altruism are (-0.351) and (-0.346), respectively. This means that, in general, subjects behaved unkindly to others and were treated badly by others. The average reciprocal altruism was (−0.37) when subjects were critical players, compared to (−0.33) when they were noncritical players. On the other hand, the average reciprocal altruism was (−0.36) when subjects were critical players, compared to (−0.33) when they were noncritical players. Hence, we can say that subjects behaved and were treated unkindly in general and that such feelings were stronger when they were critical players.

Both direct and reciprocal altruism are used to indicate the subject's kindness preferences as the final variable (v9). Due to negative reciprocal altruism, as mentioned in the previous section, the subjects' overall utilities were usually worse than their monetary payoffs.

The estimation of Probit MLEs(1) covers all observations of 2800 individual decisions, as the observations in the first round were excluded. Amongst these observations, the subjects decided to join the coalition a total of 1884 times. The result shows that past decisions, religious attitudes, the dummy variable of being a critical player, marginal benefits, past group contributions, kindness in the altruism test, and reciprocal altruism have significant effects on the decision.

The results imply that the subjects' decisions mostly followed the Nash prediction and were consistent with their past decisions. When the subjects were critical players, they were more likely to join the coalition. Moreover, larger coalition size in the past would increase the motivation to participate. It is worth noting that subjects cared about not only their own payoffs but also payoffs to others. However, the more generous they were in the altruism test, the less likely they were to join and make contributions in the public goods game. This interesting result can be illustrated with the variable of reciprocal altruism in the public goods game: when a subject was treated badly, that subject would be more likely to participate in the coalition. In other words, participation was not self-enforced but was threatened by unkind punishment.

As mentioned earlier, this experimental design set the number of critical players required to form an effective coalition. Studying the behavior of critical players can enhance our understanding of the decisions of signatories because they were essential to stabilizing the coalition internally. Probit MLE(2) and Probit MLE(4) examine the observations of first-round decisions in each treatment, whilst Probit MLE(3) and Probit MLE(5) examine the observations of decisions in the remaining rounds. In addition, the observations of critical players were shown in MLE(2) and MLE(3). Compared to the decisions in the first round and the remaining rounds, the participation rates declined from 93% to 85% and from 59% to 46% for the critical and noncritical players, respectively. This result shows that decisions did not converge to the prediction of self-interest, implying that subjects become less cooperative after learning other players' decisions. Due to the observation size in the first round, possible variables are insignificant to the decisions.

Probit MLE(3) examines the observations of critical players. A total of 85% out of the 1500 observations participated in a coalition, as the design suggested. In addition, interestingly, pro-left-wingers are more likely to participate. The critical players' decisions were consistent with their past decisions. However, if a larger coalition was formed in the previous round, those players were less likely to participate. If a subject was kind to others in the altruism test, she was less likely to participate and form a profitable coalition. The results seem irrational but can be illustrated by reciprocal altruism. The worse a subject felt about how she was treated in the prior round, the more likely she would be to participate. In other words, a subject's decision depends more on how she has been treated in the past than on her pure kindness.

It is worth noting that the participation rates in the first round were higher. Negative reciprocal altruism may provide an answer: because critical player were treated kindly when most noncritical players cooperated, they were treated unkindly most of the time. Therefore, their reactions to others became unkind due to reciprocal behavior. The more they were kind to others in the direct altruism test, the stronger the reciprocal effect to make them turn down a profitable coalition.

Having discussed the critical players, the noncritical players were assessed by estimating Probit MLE(5). These players had the free-riding incentive; however, the result shows that such incentives were rejected for nearly half of the 1200 observations. In terms of the players' preferences, individual political and religious attitudes have significant effects on the willingness to participate. Pro-left noncritical players are likely to free ride. Compared to their results when they were critical players, the pro-left-wingers played strategically by punishing and cooperating. On the other hand, the subjects with less religious belief were more likely to cooperate when they were noncritical players. The effect of reciprocal altruism was insignificant to the decisions. Neither the players' altruism attitudes in the altruism test nor their altruism attitudes in the public good game were significant to the decision. However, we claim that the noncritical players' decisions mostly depend on the marginal benefit to the total contribution. It is intuitive that higher marginal benefits brought higher incentives and led to lower participation. In contrast to the experimental evidence of (Burger & Kolstad, 2010), this study supports their earlier finding that higher marginal benefits would significantly increase the size of a coalition.

5 Conclusion

This study investigates the impact of reciprocal altruistic attitudes on individual willingness to participate in a climate coalition. The theoretical result suggests that the coalition formation could be enlarged by the participation of altruists. A particular experiment was designed to test the theory by indicating individual altruistic attitudes and individuals' willingness to join a coalition. The design assigns two player roles in the game: critical and noncritical players. The critical players have a weakly dominant strategy of joining and are essential to a profitable coalition. On the other hand, the noncritical players have a weakly dominant strategy of not joining and are dispensable to the coalition.

The experiment contains two parts: an altruism test and a public good game. In the altruism test, the result confirms the existence of altruistic preferences among 60% of the subjects. Altruistic attitudes are significantly correlated to religious attitudes, such that a stronger belief leads to a higher altruistic attitude. The incentives for participating in a coalition were examined by binary estimations through 3,000 observations in the membership games. The factors used in the binary regressions include the historical records of decisions, dummy variables of player roles, individual altruistic attitudes, age, political attitude, religious attitude, the marginal benefits of total contributions and the former coalitional formation.

The dominant strategy equilibrium design is one of the main characteristics used in this study to identify individuals' motivations. This study provides several intuitive implications: subjects' decisions were consistent and pursued higher monetary payoffs. Usually, when they were critical to the coalition, subjects followed the weakly dominant strategies of participating in a coalition. However, a kind subject in the altruism test behaved unkindly toward others in the public good game. Regarding the players' attitudes against reciprocal altruism, when they thought they had been treated badly, they were more likely to participate due to the threat of punishment. When they became noncritical, such altruistic attitudes were insignificant to their decisions. This surprising result implies that decision makers are not self-enforcing in international conventions. However, the decision process is too complicated to be captured by a single preference.

Moreover, the subjects' preferences significantly affected their decisions. The left-wingers participated more if they were critical, and they participated less when they were noncritical. This interesting result implies that they had less motivation to give up the free-riding benefit by joining a coalition. Another important aspect of self-awareness is that religionists were less likely to join a coalition, and even they were kind to others in the anonymous altruism test. Subjects with stronger religious beliefs behaved altruistically. However, this does not mean that a stronger religious attitude would lead to an altruistic decision in the interactive game. Particularly when subjects were noncritical players, a stronger religious attitude leads to a weaker motivation to participate.

Finally, this study provides policy implications by showing that self-interest remains the key factor of individual participation in climate coalitions. It is worth noting that coalition formation could be affected by reciprocal altruistic preferences. Because the decision process becomes more complicated and strategic in the interactive environment, coalition formation should be examined with a comprehensive investigation that considers other factors, including multiple individual preferences.

Tables and Figures

Table.1A. List of parameters of marginal benefit for players in Treatment 1-4

Rounds	Player 1	Player 2	Player 3	Player 4	Player 5
1-15	0.675*	0.375*	0.125	0.1	0.075
16-30	0.075	0.15*	0.25*	0.3*	0.35*
31-45	0.4*	0.65*	0.075	0.1	0.125
46-60	0.05	0.1	0.4*	0.35*	0.3*

Table.1B. List of parameters of marginal benefit for players in Treatment 5-8

Rounds	Player 1	Player 2	Player 3	Player 4	Player 5
1-15	0.075	0.1	0.45*	0.35*	0.25*
16-30	0.125	0.1	0.15	0.5*	0.55*
31-45	0.45*	0.6*	0.05	0.2	0.1
46-60	0.45*	0.25*	0.2*	0.15*	0.05

* means critical players

Table 2. The token's values for keeping (z_1), giving (z_2), the ratio of keeping to giving and the number of subjects decided to give

Round	1	2	3	4	5	6	7	8	9	10
z_1	£1	£10	£7.5	£5	£2.5	£7.5	£5	£0.5	£5	£2.5
z_2	£1	£10.5	£8	£5.5	£3	£10	£7.5	£1	£10.5	£5.5
z_1/z_2	1	0.95	0.94	0.91	0.83	0.75	0.67	0.5	0.48	0.46
Number of Giving	0	3	7	7	8	8	8	20	14	9

Round	11	12	13	14	15	16	17	18	19	20
z_1	£1	£2.5	£2.5	£0.5	£1	£1	£0.5	£1	£0.5	£0.5
z_2	£2.5	£7.5	£10	£2.5	£5.5	£7.5	£5	£10.5	£7.5	£10
z_1/z_2	0.4	0.33	0.25	0.2	0.18	0.13	0.1	0.095	0.07	0.05
Number of Giving	17	15	17	23	18	18	24	21	25	29

Table.3. OLS Estimation for Inequality-Averse Attitudes

Variable	Estimation		
Constant term	9.79 (0.63)		
Age	−0.005 (0.63)		
Politic attitude	−0.04 (0.49)		
Religious attitude	0.06* (0.07)		
Total Observation	50	R-squared	0.07
* means significant at 10% level.			

Table.4. Probit Estimations of Probability of Joining a Coalition

Variable	Probit MLE(1)	Probit MLE(2)	Probit MLE(3)
Constant term	−6.72	−14.05	2.02
(v1) Prior Decision	1.37***	2.47***	0.95***
(v2) Age	0.003	0.007	−0.00
(v3) Politic Attitude	0.04	−0.11**	0.15***
(v4) Religious Attitude	−0.05**	−0.02	−0.10***
(v5) Critical player	1.06***		
(v6) Marginal Benefit	−1.34***		−4.97***
(v7) Prior Group Contribution	0.23*	−0.50**	−0.30
(v8) Pure Altruism	−0.18 *	−0.37**	−0.02
(v9) Group Altruism	−2.35***	−3.42***	1.11
Total Observations	2800	1540	1260
Observations of Joining	1884	1308	576
LR statistic	1014.23	349.46	264.61
Note: Each cell contains coefficient. *, **, *** are significant at 10%, 5%, and 1% respectively.			

Reference

Bahn, O., Breton, M., Sbragia, L., & Zaccour, G. (2009). Stability of international environmental agreements: an illustration with asymmetrical countries. *International Transactions in Operational Research, 16*(3), 307-324.

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 878-894.

Barrett, S. (2001). International cooperation for sale. *European Economic Review, 45*(10), 1835-1850.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *The American Economic Review, 90*(1), 166-193.

Brandts, J., & Schram, A. (2001). Cooperation and noise in public goods experiments: applying the contribution function approach. *Journal of Public Economics, 79*(2), 399-427.

Bratberg, E., Tjøtta, S., & Øines, T. (2005). Do voluntary international environmental agreements work? *Journal of Environmental Economics and Management, 50*(3), 583-597.

Breton, M., Sbragia, L., & Zaccour, G. (2010). A dynamic model for international environmental agreements. *Environmental and Resource Economics, 45*(1), 25-48.

Burger, N. E., & Kolstad, C. D. (2010). *International Environmental Agreements: Theory Meets Experimental Evidence*. Retrieved from

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The quarterly journal of economics, 117*(3), 817-869.

D'Aspremont, C., Jacquemin, A., Gabszewicz, J., & Weymark, J. (1983). On the Stability of Collusive Price Leadership. *Canadian Journal of Economics, 1,* 17-25.

Dannenberg, A., Löschel, A., Paolacci, G., Reif, C., & Tavoni, A. (2015). On the provision of public goods with probabilistic and ambiguous thresholds. *Environmental and Resource Economics, 61*(3), 365-383.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior, 47*(2), 268-298.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics, 114*(3), 817-868.

Finus, M., & Rübbelke, D. T. G. (2013). Public Good Provision and Ancillary Benefits: The Case of Climate Agreements. *Environmental and Resource Economics, 56*(2), 211-226. doi:10.1007/s10640-012-9570-6

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171-178.

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

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of economic dynamics, 1*(3), 593-622.

Lin, Y.-H. (2017). *Can Individual Attitudes toward Altruism Enlarge a Climate Coalition*.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *The American Economic Review, 83*(5), 1281-1302.

Seinen, I., & Schram, A. (2006). Social status and group norms: Indirect reciprocity in a repeated helping experiment. *European Economic Review, 50*(3), 581-602.

Willinger, M., & Ziegelmeyer, A. (2001). Strength of the social dilemma in a public goods experiment: an exploration of the error hypothesis. *Experimental Economics, 4*(2), 131-144.