The Influence of Reciprocity on Individual Decisions in An Experiment

Lin, Yu-Hsuan

Department of Economics, the Catholic University of Korea

2018

The Influence of Reciprocity on Individual Decisions in An Experiment

Yu-Hsuan Lin[1]

WORKING PAPER

VERSION: JULY 2018

[**Abstract**]

This study investigates the effect of reciprocal kindness on individual decisions with experimental evidences. The literature suggested that the coalition formation could be enlarged by altruism. This study employs the reciprocity model to illustrate the reciprocal behavior in both dictator and public goods games. We found that an altruistic participant in the dictator game could be hostile to others in the public good game due to the negative reciprocal feeling. When subjects were essential to make contributions to public goods, they were more likely to cooperate if they were treated badly. In contrast, when subjects were unnecessary, the reciprocal kindness could enhance cooperative tendencies. Overall, this study reveals that the reciprocal behavior could reshape the provision of public goods.

Keywords: experimental design, reciprocity, public goods; dictator game

JEL code: C91; D64; H41

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

Acknowledgment: This study was funded by the Catholic University of Korea Research Fund.

1 Introduction

The provision of public goods has been discussed for decades. In practice, a series of discussions (such as Carraro and Siniscalco (1998); Hoel and Schneider (1997); Barrett (2001); Bratberg et al. (2005); Rubio and Ulph (2006); Finus and Rübbelke (2013) and Wu and Thill (2018)) has studied the formation of climate coalitions. As the milestone study, Barrett (1994), hypothesized that countries are self-interested and self-enforced to participate in a coalition. Most literature (such as Bahn et al. (2009); Breton et al. (2010); Finus et al. (2006)) suggested that stable coalitions achieve only little if no additional mechanism is provided in the agreement.

Smaller equilibrium coalitions have been challenged by some practical IEAs and experimental studies (Willinger and Ziegelmeyer 2001; Kosfeld et al. 2009; Burger and Kolstad 2010). They suggested that high levels of cooperation do exist in the absence of policy interventions. These studies have claimed that people are far less likely to offer a free ride and more likely to cooperate than the Nash prediction suggests. Social (or other-regarding) preferences have been proposed by recent studies (such as Charness and Rabin (2002); Hadjiyiannis et al. (2012) and Dannenberg et al. (2015)) to address this knowledge gap.

Whilst 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 hostile 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 and Kirchsteiger 2004), and the quest for efficiency gains through (Brandts and Schram 2001). There are also overviews of these (Seinen and Schram 2006). The reciprocal model has been proposed to illustrate the experimental results, such as ultimatum bargaining game (Dickinson 2000) and public goods games (Bardsley and Moffatt 2007). In this paper, we intend to contribute to discussions of the motivations underlying reciprocal altruism and build a bridge to connect experimental evidences.

In a previous study, Lin (2017) considered the impact of an unidirectional individual altruistic preference for the formation of a climate coalition. The study provided a particular design of dominant strategy equilibrium to predict individual behaviors and coalition formation. However, the experimental result showed that a player could be kind in a unidirectional dictator game but be hostile in an interactive public-good game. It means unidirectional altruistic model was unable to predict decisions in an interactive game. In order to understand the cooperative behavior of individuals, this study employs the model of reciprocity.

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. 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 examines individual reciprocal preferences by employing Rabin (1993)'s framework of reciprocity and Lin (2017)'s experimental evidence. The reciprocal model considers a player's own payoff, the player's perception of others' payoffs, and others' perception of the player's payoff. The experimental design considers the public goods game but limited in a scope of dominant strategy equilibrium. It provides the primary strength of investigating individual incentives for participating in a coalition. If there were more than one stable coalition, the individual decisions were difficult to be predicted. A dominant strategy equilibrium, therefore, provides a suitable environment in which to observe individual decisions when every player had an optimal strategy to choose.

With the reciprocal preference, the coalition formation could be reshaped and different from the dominant strategy equilibrium. Due to the interactive perceptions of players' payoffs, the coalition size might become smaller or larger than the self-interested dominant strategy equilibrium. 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 study consists of five sections: section 2 describes the model of reciprocity.

Section 3 illustrates the experimental design. Section 4 analyses the reciprocal behavior with experimental evidence. The final section concludes.

2 Model of Reciprocity

Consider a reciprocity model with N players, following Rabin (1993), a strategy set S_i for player i where $i \in [1, N]$. The material payoff of a player π_i depends on the strategies chosen by players i and others ($S_i \times S_{-i}$). We use the following notation: $a_i \in S_i$ represents the actual strategies chosen by the player; $b_i \in S_i$ represents other non-i player's belief about which strategy player i is choosing; $c_i \in S_i$ represents player i's belief about what other players believe player i's strategy is.

We assume that player i's subjective expected welfare depends on three kindness preferences: (i) player i's own interest, (ii) other players' interests and (iii) the player i's interest influenced by other players' strategies. A player i chooses a strategy to maximize her expected welfare, which incorporates both her material utility and the players' shared notion of fairness:

$$(1) \qquad u_i = u_i^s(\pi_i) + u_i^r(\pi_i, \pi_{-i})[1 + u_i^a(\pi_i, \pi_{-i})]$$

The first term u_i^s is pure self-interest, the second term contains reciprocal kindness, u_i^r, and straight kindness, u_i^a. Both straight kindness and reciprocal kindness are in the range of -1 and $1/2$.

Reciprocal kindness indicates how player i experiences other players' kindness, while straight kindness indicates player i's kindness to other players. The impact of reciprocal kindness in welfare depends on a player's feeling for others. If the player feels being treated badly, her overall welfare will be lower than her monetary payoff. On the other hand, straight kindness plays the role of strength of feeling. If a player is straightly hostile, she cares less about others' decisions. But if a player is straight generous, her welfare depends on the communication of kindness: if she is treated kindly, her welfare is higher; if she is treated badly, her welfare is lower than the monetary payoff. These types of kindness are defined as follows.

Definition 1: Reciprocal kindness defines player $i's$ belief about how kind other players are being to her

$$(2) \qquad u_{-i}^r(b_{-i}, c_i) \equiv \frac{\pi_i(c_i, b_{-i}) - \pi_i^e(c_{-i})}{\pi_i^h(c_i) - \pi_i^{min}(c_i)}$$

if $\pi_i^h(c_i\ ) - \pi_i^{min}(c_i\ ) = 0$, then $u_i^r(b_{-i}, c_i\ ) = 0$.

Player i's reciprocal kindness consists of her payoffs: $\pi_i\ (c_i\ , b_{-i})$ is the payoff that i chooses what others believe and others choose what i believes. $\pi_i^h(c_i\ )$ and $\pi_i^l(c_i\ )$ are the highest and lowest payoffs in the set of all feasible Pareto-efficient payoffs respectively. $\pi_i^e(c_{-i})$ is the *equitable payoff* defined as the average of the highest and lowest payoffs, $[\pi_i^h(c_i\ ) + \pi_i^l(c_i\ )]/2$; and $\pi_i^{min}(c_i\ )$ is the worst possible payoff for player i in the set of all possible payoffs.

Definition 2: Straight kindness defines player i's kindness to other non-i players

(3) $$u_i^a(b_{-i}, a_i\ ) \equiv \frac{\pi_{-i}(b_{-i}, a_i\ ) - \pi_{-i}^e(b_{-i}\ )}{\pi_{-i}^h(b_{-i},) - \pi_{-i}^{min}(b_{-i})}$$

if $\pi_{-i}^h(b_{-i},) - \pi_{-i}^{min}(b_{-i}) = 0$, then $u_i^a(b_{-i}, a_i\ ) = 0$.

On the other hand, Player i's straight kindness consists of the payoffs of other players: $\pi_{-i}(b_{-i},\ a_i\ )$ is a non-i player's payoff that what i chooses and what i believes others' would do. $\pi_{-i}^h(b_{-i},\ a_i\ )$ and $\pi_{-i}^l(b_{-i},)$ are the highest and lowest payoffs which a non-i player could get in the set of all feasible Pareto-efficient payoffs respectively. The equitable payoff, $\pi_{-i}^e(b_{-i}\ )$, is the average of the highest and lowest payoffs; and and $\pi_i^{min}(c_i\ )$ is the worst possible payoff for a non-i player in the set of all possible payoffs.

In practice, we assume that beliefs b_i and c_i are player i's actual behavior in the past. In other words, what a player would believe and would like the others to believe are based on the past decisions. Thus $\pi_i\ (c_i\ , b_{-i})$ is player i's payoff in the past round, while $\pi_{-i}(b_{-i},\ a_i\ )$ is a non-i player's payoff given i's present decision and others' past decisions.

3 Experimental Design

This study employs the reciprocal model to illustrate individual behaviour in a laboratory experiment of Lin (2017). There were fifty students with multi-cultural and multi-disciplinary backgrounds were invited. The experiment consisted of two parts: the first part is a dictator game which examined individual altruistic attitudes, and the second part is a public goods game which measured interactive social preferences.

In the dictator game, 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.

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 the dictator game, 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 straight kindness level is either -0.5 (keep) or 0.5 (give).

Turning now to the experiment on public good provision, subjects were randomly assigned to groups of 5 persons for the whole session, which was conducted anonymously. 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. Individuals decided whether to contribute to the group by joining or not to contribute by not-joining. 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 public good provision.

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.

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.

In an effective coalition, a signatory's payoff is the summation of all signatories' marginal benefit subtract the standard abatement cost whilst a nonsignatory's payoff is the product of her individual benefit and the coalition size. When an effective coalition collapses, all players get nothing. It is worth noting that, this mode focuses on a dominant strategy equilibrium which players are categorized into two types: critical and non-critical to an effective coalition. Critical players have a dominant strategy to cooperate whilst non-critical players have a dominant strategy to free-ride. Hence, a critical player could reach the highest payoff when all players cooperate, while a non-critical player reach her highest payoff when she is the only free-rider. The lowest Pareto-efficient payoff is the payoff in the dominant strategy equilibrium which critical players participate for the shared coalition payoff and non-critical players do not participate for the free-riding benefit. As defined, the equitable payoff $\pi_{-i}^{e}(b_{-i})$ is the average of the highest and lowest payoffs. Besides, the worst possible payoff for others $\pi_{-i}^{min}(b_{-i})$ would be 0.

Notice that the signs of reciprocal and straight kindness depend on the numerators since the denominators are positive number. When the player earns less than the equitable payoff, the player would think other players are hostile to her. When other player earns less than the equitable payoff, the player is hostile to others.

Consider an example when all players participate in a coalition, the collective payoff reaches the highest level. A critical player has positive straight and reciprocal kindness. It means the player's welfare is more than her monetary payoff in a grand coalition. In contrast, due to the unreached free-riding benefit, a non-critical player feels other players are being hostile to her. The player's welfare would be lower than her monetary payoff. Thus, non-critical players would be unwilling to coordinate with others.

Take the example of the dominant strategy equilibrium, a non-critical player is hostile toward other players because she is unable to reach the highest free-riding benefit. Her welfare would be less than her material payoff. On the other hand, a

critical player behaves kindly to others but feels non-critical players are mean to her. Such hostility may lead other players to undertake costly punishment by the critical player. The consequence of dominant strategy equilibrium would become unstable internally.

When a critical player decides to take revenge with non-cooperative behavior, the coalition becomes ineffective and everyone earns nothing. It means all possible responses yield other players the same payoff. Therefore, there is no issue of kindness. This situation would be changed when the critical player believes that other players would behave cooperatively. The player would participate when she believes the coalition could be larger than the dominant strategy equilibrium.

Together these results provide important insights into the unstable coalition formation, due to players' beliefs and reciprocal behaviour. In a grand coalition, non-critical players might feel hostile for the unreached free-riding benefit. In the dominant strategy equilibrium, critical players might feel hostile toward the free-riders. Thus the coalition formation could be reshaped by their beliefs and preferences.

4 Analyses with experimental evidence

Turning now to the experimental evidence on the dictator game and the public goods game, the observation of subjects' decisions illustrates the straight and reciprocal kindness.

In the dictator game, 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.

As mentioned earlier, the constrained design limited the observation on the subjects' straight kindness attitudes in the dictator game. The subject's average straight kindness level in 20 rounds is (-0.21), which implies that subjects were hostile to others in general. In general, an increasing trend is noticed, such that subjects became 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, subjects were more likely to behave kindly by giving it up.

Table 3 shows the OLS estimation of straight kindness attitude. The dependent

variable is the individual's average straight kindness 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.

In the second part of the experiment, subjects played a public goods game by deciding whether or not to participate in a coalition. The result shows that 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 dominant strategy equilibrium in only 112 rounds. The coalition was usually neither stable nor convergent to a particular formation. With the same treatments, the coalition formation varied in different groups. This implies that, even with the design of dominant strategy equilibrium, a stable formation was unreachable.

Compared to the individual decisions in the first round, the participation rates in the remaining rounds declined from 93% to 85% and from 59% to 46% for the critical and noncritical players, respectively. The result shows that individual decisions did not converge to the prediction of self-interest, implying that subjects become less cooperative after learning other players' decisions.

In order to understand 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 year the subjects were born, (v2) political attitudes from left to right, (v3) religious attitudes from atheist to religionist, (v4) the dummy variable of being critical players, (v5) the marginal benefit of total contribution, (v6) the group contribution in the previous round, (v7) straight kindness attitude in the dictator game and (v8) the player's shared notion of fairness in the coalition game.

The last two factors are related to the reciprocal preference. Firstly, we use the subject's average straight kindness level in 20 rounds in the dictator game. Secondly, we used the subjects' shared notion of fairness which incorporates straight kindness and reciprocal kindness in the coalition game. As mentioned earlier, the decisions made in the prior round are used to indicate players' beliefs. 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 reciprocal kindness attitudes could be indicated as the average of her attitudes to other four players in the group.

The estimations used 2,800 observations (due to the exclusion of the first observation in every treatment) in the coalition game. A correlation coefficient was computed to assess the relationship between an individual's straight and reciprocal kindness. There was a strong and positive correlation (0.84) between straight and reciprocal altruism. The mean values of direct and reciprocal altruism are (-0.351) and (-0.346), respectively. This means that, in general, subjects were hostile 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 badly in general and that such feelings were stronger when they were critical players.

The estimation of Probit MLEs(1) covers all observations of 2800 individual decisions. Amongst these observations, the subjects decided to join the coalition a total of 1884 times. The result shows that religious attitudes, the dummy variable of being a critical player, marginal benefits, past group contributions, straight kindness attitude in the dictator game, and the player's shared notion of fairness in the coalition game. When the subjects were critical players, they were more likely to participate in the coalition. The intuition behind the result is that they could earn monetary reward by doing so. Moreover, larger coalition size in the past would increase the motivation to participate.

Due to negative reciprocal altruism, as mentioned in the previous section, the subjects' overall utilities were usually worse than their monetary payoffs. 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 kindness 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 only self-enforced but also complied with a hostile 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) examines the observations when they were critical players, whilst Probit MLE(3) examine the observations when they were non-critical players.

Probit MLE(2) examines the observations of critical players. If a larger coalition was formed in the previous round, those players were more likely to participate. If a subject was kind to others in the dictator game, she was less likely to participate and form a profitable coalition. The result seems 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, how she has been treated in the past could alter the decisions. It is worth noting that the participation rates in the first round were higher. Negative reciprocal altruism may provide an answer: in a coalition larger than the dominant strategy equilibrium, critical players were treated kindly when most noncritical players cooperated while non-critical players were treated badly. Therefore, their reactions to others became unkind due to reciprocal behavior. The more they were kind to others in the 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(3). These players had the free-riding incentive; however, the result shows that such incentives were rejected for nearly half of the 1200 observations. Interestingly, individual political and religious attitudes have significant but different effects on critical and non-critical players. Pro-left noncritical players are likely to free ride. Compared to their results when they were critical players, the pro-left-wingers were more likely to participate. In terms of the players' religious preferences, the subjects with less religious belief were more likely to cooperate, particularly when they were non-critical players.

The effect of reciprocal kindness was significant but different to non-critical players, compared to critical ones. The higher shared notion of fairness, non-critical

players were more likely to cooperate. The result implies that the hostility encouraged players giving up the free-riding benefits. However, higher marginal benefits brought higher incentives, the free-riding incentive could still led to lower participation. In contrast to the experimental evidence of (Burger and Kolstad 2010), this study against their earlier finding that higher marginal benefits would significantly increase the size of a coalition.

5 Conclusion

This study investigates the impact of reciprocal kindness attitudes on individual decisions with experimental evidences. The theoretical result suggests that, depends on the reciprocal preference, individual welfare could be more or less than the monetary payoff. The reciprocity model illustrates the individual decisions in a particular experiment which consists of a dictator game and a public goods game. The public goods game was designed in a dominant strategy equilibrium. The design made players becoming either critical or non-critical to an effective coalition. 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 Straight Kindness 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	−2.69	−6.21	7.27
(v1) Age	0.001	0.003	−0.003
(v2) Politic Attitude	0.08 **	−0.12 **	0.20 ***
(v3) Religious Attitude	−0.05 **	0.01	−0.12 ***
(v4) Critical player	1.17 ***		
(v5) Marginal Benefit			−631 ***
(v6) Prior Group Contribution	0.78 ***	1.29 ***	−0.30
(v7) Straight Kindness in Dictator Game	−0.21 **	−0.27 **	−0.06
(v8) Shared Notion of Fairness	−0.90 ***	−2.13 ***	3.81 ***
Total Observations	2800	1540	1260
Observations of Joining	1884	1308	576
LR statistic	570.74	101.66	130.83
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.

Bardsley, N., & Moffatt, P. G. (2007). The Experimetrics of Public Goods: Inferring Motivations from Contributions. [journal article]. *Theory and Decision, 62*(2), 161-193, doi:10.1007/s11238-006-9013-3.

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, doi:https://doi.org/10.1016/j.jeem.2005.03.002.
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. University of California at Santa Barbara.
Carraro, C., & Siniscalco, D. (1998). International Institutions and Environmental Policy: International environmental agreements: Incentives and political economy1. *European economic review, 42*(3-5), 561-572.
Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics, 117*(3), 817-869.
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.
Dickinson, D. L. (2000). Ultimatum decision-making: A test of reciprocal kindness. [journal article]. *Theory and Decision, 48*(2), 151-177, doi:10.1023/a:1005274316908.
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. [journal article]. *Environmental and Resource economics, 56*(2), 211-226, doi:10.1007/s10640-012-9570-6.
Finus, M., Van Ierland, E., & Dellink, R. (2006). Stability of climate coalitions in a cartel formation game. *Economics of Governance, 7*(3), 271-291.
Hadjiyiannis, C., İriş, D., & Tabakis, C. (2012). International environmental cooperation under fairness and reciprocity. *The BE Journal of Economic Analysis & Policy, 12*(1).
Hoel, M., & Schneider, K. (1997). Incentives to participate in an international environmental agreement. *Environmental and Resource economics, 9*(2), 153-170.
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.
Rubio, S. J., & Ulph, A. (2006). Self-enforcing international environmental agreements revisited. *Oxford economic papers, 58*(2), 233-263.
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.

Wu, J., & Thill, J.-C. (2018). Climate change coalition formation and equilibrium strategies in mitigation games in the post-Kyoto Era. *International Environmental Agreements: Politics, Law and Economics*, 1-26.