How Social Preferences Influence the Stability of a Climate Coalition

Lin, Yu-Hsuan

Department of Economics, the Catholic University of Korea

March 2018

How Social Preferences Influence the Stability of a Climate Coalition

Yu-Hsuan Lin[1]

WORKING PAPER

VERSION: March 2018

[Abstract]
This study examines the impact of social preferences on the individual incentives of participating in climate coalitions with laboratory experimental evidences. The theoretical result suggests that, when a player was inequality-neutral, a dominant strategy equilibrium could exist. However, individuals with social preference may lead a vacillated coalition formation. Joining or not joining depend on the player was critical or non-critical to an effective coalition respectively. The laboratory experimental result shows that players were inequality-averse and the coalition was usually larger than the equilibrium size but unstable. The inequality-averse attitudes have significantly positive impact on the incentives of participation. Particularly, when they are non-critical players, egalitarians are likely to give up the free riding benefit by joining a coalition. Our findings help to understand the climate coalition formation.

[Keywords]: international environmental agreements; social preference; inequality-aversion; experimental design; climate coalition

[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 work was supported by the Catholic University of Korea Research Fund.

1. Introduction

Since Barrett (1994), a large number of studies (such as Bahn et al. (2009), Barrett (2001), Bratberg, Tjøtta, and Øines (2005), and Eyckmans and Finus (2006)) have explored the formation of international environmental agreements (IEAs) on climate change. Most theoretical predictions on IEAs predicted a rather low number of signatories to a stable self-enforcing coalition (Grüning and Peters 2010). This is in sharp contrast to empirical evidence. Recent experimental studies (Burger and Kolstad 2010, Kosfeld, Okada, and Riedl 2006, McEvoy et al. 2014) have pointed out that actual coalition formations are usually larger than theoretical predictions. A growing number of studies have proposed that this challenge is due to the fundamental assumption of rational self-interest (Willinger and Ziegelmeyer 2001).

The assumption of self-interest, has been widely employed in the majority of studies of IEAs, requires that rational agents pursue higher payoffs (Breton, Sbragia, and Zaccour 2010). Such assumption was not enough to explain individual decision makers' behaviours in a climate convention. Several studies (Lange 2006, Grüning and Peters 2010, Charness and Rabin 2002, Dannenberg et al. 2015) had suggested that the role of social preferences (also known as other-regarding preferences) should be considered to address this limitation.

Several survey studies had also identified individuals' social values and norms on preferences toward climate policies (Alló and Loureiro 2014, Domínguez Arcos, Labandeira Villot, and Loureiro García 2011, Hanemann, Labandeira, and Loureiro 2011, Svenningsen and Thorsen 2017, Svenningsen 2017). To reach higher environmental standard for climate change mitigation, people with social values were willing to contribute more.

As experimental economics provided some evidence for more complex human behaviour, extending the theory of IEAs to a broader class of preferences is clearly promising. Most of them, such as Kosfeld, Okada, and Riedl (2006) and Grüning and Peters (2010), suggested that countries' preferences incorporate justice and fairness could promote abatement and enlarge the coalition size. On the other hand, Kolstad (2014) argued that such social preference reduced the equilibrium size of a coalition of agents formed to provide the public good.

Although the influence of social preferences on coalition formation has be discussed, the influence on individual incentives for participating in a coalition has not yet been properly explored. A design with multiple equilibria, where agents have no dominant strategy, would limit the prediction of individuals' decisions. Hence, this study answers two questions: Does the concern about payoffs of others affect to individuals' decisions? How do such individual social preferences the coalition formation?

This study investigates the influence of individual social preferences in a climate coalition with experimental evidences. A theoretical model with a dominant strategy equilibrium was built to predict individual behaviours. If players were inequality-neutral, based on their weakly dominant strategies, they could be either critical or non-critical. However, if players were inequality-averse, such equilibrium may not exist. There would be three possible outcomes: first, if players do not concern much about fairness, the outcome remains the same. Second, when critical players were inequality-averse, no stable coalition exists. Third, when non-critical players were inequality-averse, an effective coalition would be larger than the dominant strategy equilibrium. Such theoretical predictions were tested by a two-part laboratory experiment: the first part identifies individual inequality-averse attitudes by taking either a certain fair payoff or an all-or-nothing payoff. The second part was a membership game by asking subjects whether or not to join a coalition.

There are two primary strengths in this study. First, individual decisions could be predictable in the model. This experiment design would provide a suitable environment to observe individual decisions when every player had a weakly dominant strategy. Second, in order to examine the impact of social preference on coalition formation, the theoretical prediction was compared with the actual experimental outcome. The design provides detailed observations on the process of individual decision and coalition formation.

The outline of the study is as follows: In the next section, a benchmark model and an inequality-averse model were built. Section 3 introduces two experiments that were based on the theory built in Section 2. Then, Section 4 reports the experimental results and implications. The final section concludes.

2. The Models

This study builds a simple climate coalition game: players decide their membership. Signatories would do fully abate whilst nonsignatories would do fully pollute. The game considers two scenarios: the first scenario assumes that countries are self-interested and the second scenario assumes that countries have social preferences.

Suppose N countries with different marginal benefits of the total abatement considered participation in a climate coalition, and then n countries joined while the rest did not. Each country has a heterogeneous marginal benefit, γ , in the range of 0 to 1 and an identical standardized unit cost of abatement.

Countries self-enforced to join an IEA and their payoffs depended on their membership status. In a profitable n-member coalition, the payoff of a non-signatory j (π_j) is $\gamma_j X$, where γ_j is its marginal benefit rate, and X is the total abatement from signatories. On the other hand, all of the signatories choose same abatement level in order to maximise the coalition payoff, then share the coalition payoff equally. The coalition payoff (Π_s) is the joint payoffs of n members ($\pi_i, \forall i = 1, \dots, n$). The uniform abatement can be denoted as x_s. The coalition payoff is represented as $\Pi_s = \sum_{i=1}^{n} \pi_i = \sum_{i=1}^{n} [(-x_s) + \gamma_i X]$.

2.1 Benchmark model with self-interest preference

In the scenario of self-interest, countries concern only their own payoffs. Nonsignatories choose pollute and signatories choose abate. Hence the payoffs of a nonsignatory j and a signatory i are presented respectively as

$$\pi_j = \gamma_j X \quad \text{(Eq.1)}$$

$$\pi_i = (-1) + \sum_{i=1}^{n} \gamma_i \quad \text{(Eq.2)}$$

In the membership game, countries were asked to simultaneously decide whether to participate in a coalition or not. Following d'Aspremont et al. (1983), when countries are self-interested, a stable coalition exists when both the internal and the external constraints are satisfied as follows:

$$u_i\ (n^*) > u_j\ (n^* - 1) \quad \text{(Eq.3)}$$

$$u_j\ (n^*) > u_i\ (n^* + 1) \quad \text{(Eq.4)}$$

A country's welfare function, $u(n)$, is equal to the country's payoff which depends on the number of signatories (n) and its membership status indicated by the subscript (i means signatory; j means nonsignatory). The internal constraint (Eq.3) denotes that a signatory have no incentive to leave the n^*-member coalition where n^* is the stable number to maintain the coalition. The external constraint (Eq.4) indicates that a non-signatory have no incentives to participate in a coalition as the $(n^* + 1)$ th member. When both (Eq.3) and (Eq.4) are satisfied, a stable coalition would exist. Taking examples of experimental studies, such as Kosfeld, Okada, and Riedl (2006) and Burger and Kolstad (2010), there existed several stable coalition combinations. Due to no dominant strategy for countries, these studies failed to foresee individual decisions in the membership game. This study, in order to have better prediction on

individual decisions, focus on the cases of dominant strategy equilibrium. To reach such unique equilibrium, an additional condition should be satisfied:

$$\gamma_{n^*} > \sum_{j=n^*+1}^{N} \gamma_j \qquad \text{(Eq.5)}$$

The condition categorises countries into two groups: *critical* and *non-critical* countries. Critical countries, with large marginal benefits, are essential to a profitable coalition. Therefore, the weakly dominant strategy of critical players is participating in a coalition. Non-critical countries, with small marginal benefits, can contribute the coalition but not necessary. The weakly dominant strategy of non-critical players is not to participate. The condition (Eq.5) implied that any critical country cannot be replaced by all of the non-critical countries. The condition ensures that the coalition is the only stable combination to be profitable. In other words, critical countries would participate in a coalition because they were necessary members and non-critical countries would not participate because they could take advantage from free-riding. While we acknowledge this is indeed a strong condition, in order to identify the individual incentives to participate in the coalition, this condition provided better observation of the individual decisions in the membership game.

2.1 Inequality-averse preference in a coalition game

We now turning to discuss the scenario of inequality-averse preferences. Following Fehr and Schmidt (1999), the magnitude of inequality-aversion indicates the level of dislike for unfair outcomes. With this concept, a country k ($k \in [1, N]$) has a different welfare function

$$u'_k = \pi_k - \frac{\alpha_k}{N-1}\sum_{k'} \max\left(\pi_{k'} - \pi_k, 0\right) - \frac{\beta_k}{N-1}\sum_{k'} \max\left(\pi_k - \pi_{k'}, 0\right) \qquad \text{(Eq.6)}$$

where k' is any other country except k. The first term is the payoff of country k. The second and third terms represent the average payoff gap from the other country k' with the disadvantage-loss parameter α_k and advantage-loss parameter β_k, respectively. Both of the parameters presented an inequality-averse magnitude of k and were between 0 (inequality-neutral) and 1 (strongly inequality-averse).

As previously mentioned, when countries were self-interested and the constraints of (Eq.3), (Eq.4), and (Eq.5) were satisfied, a stable n^*-member coalition existed. When inequality-aversion is taken into account, the coalition formation depends on the individual inequality-averse magnitudes. The coalition formation could become either a stable n^*-member coalition, unstable, or a stable coalition larger than n^*.

When all of the countries were inequality-neutral or weakly inequality-averse, a stable n^*-member coalition exists as (Eq.3), (Eq.4) and (Eq.5) stand. When any critical country was strongly inequality-averse, the country feels disadvantaged from the payoff gap between players. (Eq.3) would be violated by the absence of such country from the effective coalition. Nevertheless, the country has the incentive to participate if everyone yields nothing from a collapsed coalition. Therefore, the coalition formation became unstable. In the last circumstance, when any non-critical country was strongly inequality-averse, that country could participate to mitigate the payoff gap. Therefore, (Eq.4) would be violated and the coalition size is expanded and larger than n^*.

Intuitively, there were a number of effects with inequality-aversion. First, egalitarianism reduces the individual welfare when the payoffs among the countries were not equal. A coalition could be enlarged by a non-critical egalitarian country, when it sought for smaller advantage loss. Second, the transfer mechanism where signatories equally shared the coalition payoff could minimise the payoff gap among the countries. However, except for a grand coalition, signatories always suffered the disadvantage loss from non-signatories. An expanding IEA tended to exacerbate the payoff gap between the signatories and the non-signatories. Egalitarian signatories could punish free-riders behaviour by turning down a profitable IEA. In other words, the effects of inequality-aversion could shape the stability and the formation of IEAs both internally and externally.

3. Experiment Design and Procedure

The experiment was conducted at the Centre for Experimental Economics (EXEC) laboratory at the University of York (UK) and programmed with z-Tree (Fischbacher 2007). Fifty subjects were invited through the Online Recruitment System (ORSEE) (Greiner 2004). They were students from different countries and studied various disciplines. In order to understand the coalition formation, we mimicked the diversity in the real world where decision makers have different nationalities and multidisciplinary knowledge in this experiment.

In order to ensure data quality, the subjects had to comprehend the rules of the game as much as possible. They were not allowed to exchange information and no conversation was allowed (except for asking the experimenter to clarify the questions) during the experiment. The experimenter introduced the rules and gave the participants time to read through the instructions thoroughly and to accomplish the

controlled questions. At the beginning of each part of the experiment, four control questions were asked in order to test the subjects' understanding. A new part would only start if all of the subjects had answered all of the control questions correctly.

A pre-experimental questionnaire was conducted in order to gather demographic information, including the subjects' degree disciplines, age (the year they were born), ethnicity, political orientation, and their level of belief in a religion. Another two questions collected information about their self-evaluated preferences. The question regarding religion identified the subjects' belief attitude on a scale ranging from 1 (not religious at all) to 5 (extremely religious). The distribution of the level of religious attitude showed that most subjects considered themselves as mild belief. The last question aimed to indicate the subjects' political preference (level one indicates left, level two indicates centre-left, level three indicates neutral, level four indicates centre-right, and level 5 indicates right). The distribution showed that most respondents were pro-left wingers.

The experiment was comprised of two parts and its procedures were designed as follows.

3.1 An inequality-averse preference test

In this test, we aimed to examine the individuals' attitudes towards inequality-aversion. In order to extract information from a purified environment, the subjects were paired without knowing their partners or their partners' decisions. Each subject had two roles: dictator and receiver. A receiver passively earned allowance from the dictator's decision. A dictator, on the other hand, decided to share a £5 allowance with his/her receiver. There were two ways to share as shown in Table.3.1. Option 1 shared the allowance equally, while option 2 allocated the allowance unjustly with an all-or-nothing allocation at a certain probability.

Table.1. Inequality-aversion Test

Round	Option 1	Option 2
1	(£2.5, £2.5) for sure	(£0, £5) with probability 0%; (£5, £0) with probability 100%
2	(£2.5, £2.5) for sure	(£0, £5) with probability 10%; (£5, £0) with probability 90%
3	(£2.5, £2.5) for sure	(£0, £5) with probability 20%; (£5, £0) with probability 80%
4	(£2.5, £2.5) for sure	(£0, £5) with probability 30%; (£5, £0) with probability 70%
5	(£2.5, £2.5) for sure	(£0, £5) with probability 40%; (£5, £0) with probability 60%
6	(£2.5, £2.5) for sure	(£0, £5) with probability 50%; (£5, £0) with probability 50%
7	(£2.5, £2.5) for sure	(£0, £5) with probability 60%; (£5, £0) with probability 40%
8	(£2.5, £2.5) for sure	(£0, £5) with probability 70%; (£5, £0) with probability 30%
9	(£2.5, £2.5) for sure	(£0, £5) with probability 80%; (£5, £0) with probability 20%
10	(£2.5, £2.5) for sure	(£0, £5) with probability 90%; (£5, £0) with probability 10%
11	(£2.5, £2.5) for sure	(£0, £5) with probability 100%; (£5, £0) with probability 0%

Option 1 was a fair allocation where the dictator faces no unfair loss. On the other hand, the all-or-nothing allocation in option 2 indicated two extreme cases. As described in (3.6), an inequality-averse agent considered both advantage-loss and disadvantage-loss. The range of the two extremities could be normalised. The range between two extreme unfair outcomes was normalised, so both the advantage- and disadvantage-losses could be merged as one inequality-averse indicator. Although a subject might suffer more from disadvantage than advantage, two reasons supported this technique. In practice, it is not easy to find a subject's preference without standardising the unit of the utility. In the literature, the experimental evidence showed that the disadvantage factor was not necessarily smaller than the advantage factor (Dannenberg et al. 2012, Yang, Onderstal, and Schram 2016). In round 1, the all-or-nothing allocation would be taken by a rational subject because the outcome was definitely better than that from option 1. By contrast, in the final round, the outcome for the fair allocation was better than that for the all-or-nothing allocation. For each subject with a consistent preference, there existed a point with a certain probability where the subject would switch from the all-or-nothing allocation to the fair allocation. The switch point indicated the individuals' attitudes toward inequality-aversion.

When the subjects were inequality-neutral (or self-interested), then their welfares were the same to their monetary payoffs. In other words, they would switch when the expected outcome of all-or-nothing allocation was equal to that of fair allocation. When subjects were inequality-averse, their utilities were lower than their monetary payoffs. They were more likely to take an equal allocation in order to avoid extremely unfair consequences. Inequality acceptors, which never chose an equal allocation, could be possible, but they were uncommon in reality (as seen in the experimental results later). They could be observed in this experimental design. Therefore, we excluded those inequality acceptors from our analyses, similar to (Fehr and Schmidt 1999).

It is important to bear in mind that this test could be characterised by strategic uncertainty due to the fact that a series of probabilities were involved. The subjects' risk attitudes might have been involved in their decisions. In other words, it might have been difficult to distinguish the risk aversion and inequality aversion in this study. This issue might be avoided by employing two separate games in order to indicate the attitudes toward disadvantage- and advantage-aversion, such as those developed by Blanco, Engelmann, and Normann (2011) and Yang, Onderstal, and Schram (2012). However, this study was superior for two reasons. First, the two-games created another bigger issue in that the measurement of two attitudes might have been biased. Second, there was a significant positive correlation between the inequality-aversion and risk-aversion (Carlsson, Daruvala, and Johansson-Stenman 2005, Kroll and Davidovitz 2003). It was unnecessary to distinguish the inequality-aversion from the risk-aversion.

3.2 Coalition game experiment

This section mimic the international environmental negotiation by building a public goods game. The subjects were randomly assigned different roles in a group of five anonymous persons for the entire session. As described in Equations (1) and (2), the payoffs depended on the marginal benefit of the total abatement. In this study, we built eight treatments of various marginal benefits. Each group played four treatments for a session. Each treatment had 2 to 4 critical players whilst the rest played a role of non-critical. As explained earlier, based on the assumption of self-interest, the unique-equilibrium design could help to identify individual decisions. As illustrated earlier, critical players were essential for a profitable coalition, while non-critical 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, non-critical 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. The payoff depended on the given parameters and the coalition formation. For any unprofitable coalition, all of the subjects in the group gained nothing in return. The possible payoffs for the subjects ranged from £0 up to £24.

4. Experimental results

The results for the inequality-averse test demonstrated that 31 out of the 50 subjects had clearly switched from the all-or-nothing allocation to the fair allocation. In particular, 2 subjects stuck with the fair allocation for the entire session. Their behaviours indicated their individual attitude toward inequality-aversion.

Table.3. Number of Fair Allocation taken

Round	1	2	3	4	5	6	7	8	9	10	11
Number for Fair Allocation Taken	10	11	8	23	33	35	48	48	47	48	46

Table 3 presents the number of fair allocations taken in each round. Initially, most of the subjects preferred the all-or-nothing allocation. Then, their decisions switched to the fair allocation. When the expected payoff of all-or-nothing became lower than that of the fair allocation, it is unsurprising that almost everyone took the fair allocation.

Regarding the coalition formation in the membership game, profitable 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 profitable coalitions in only 12 rounds.

In this study, we predicted the individual incentives of participating in a coalition by employing the subjects' inequality-averse attitudes and historical decisions in the membership game. On the other hand, benchmark self-interested predictions were

built by employing only the historical data in the membership game. The inequality-averse predictions matched the actual decisions by 1,838 over 2,800 observations (65.6%). In the sample of 1,540 observations of critical subjects, the inequality-averse predictions matched the actual outcome for 77.2% of the observations. On the other hand, the inequality-averse predictions matched for 51.5% of the 1,260 observations. With individual inequality-averse predictions, the predicted coalition formation was shown as unstable, but it was usually larger than the actual formation.

Table.5. Probit Estimations of Probability of Joining a Coalition

Variable	Probit MLE(1)	Probit MLE(2)	Probit MLE(3)	Probit MLE(4)	Probit MLE(5)
Constant term	8.98 (12.36)	0.52*** (0.16)	-1.69 (20.99)	-0.49*** (0.15)	16.57 (18.65)
Past Decision	1.23*** (0.07)		1.34*** (0.13)		1.00*** (0.09)
Inequality-Averse Attitude	0.04** (0.02)	0.07*** (0.02)	0.05** (0.02)	0.05*** (0.02)	0.03 (0.09)
Age	-0.004 (0.006)		0.001 (0.01)		-0.008 (0.009)
Politic Attitude	0.05 (0.03)		-0.13** (0.05)		0.22*** (0.05)
Religious Attitude	-0.05** (0.02)		0.03 (0.03)		-0.17*** (0.03)
Critical player	0.76*** (0.06)				
Marginal Benefit					-6.37*** (1.11)
Past Group Contribution	-0.20* (0.12)		-0.26 (0.21)		-0.34** (0.15)
Total Observations	2,520	1,500	1,400	1,200	1,120
Observations of Joining	1,692	1,279	1,185	555	507
Log Likelihood	-1219.83	-621.21	-513.39	-824.28	-771.30

Note: Each cell contains coefficient and standard error in parenthesis. *, **, *** are significant at 10%, 5%, and 1% respectively.

Turning now to the factors that might have affected the individual decisions, the maximum likelihood estimation of the binary probit regressions (Probit MLE) were employed as shown in Table 3.5. The variables included the decision made during the previous round, the times of taking the fair allocation, the year the subjects were born, the political attitudes from left to right, the religious attitudes from atheist to

religious, the dummy variable of being critical players, the marginal benefit of the total contribution, and the group contribution in the previous round.

Although the experimental design allowed for the existence of inequality acceptors, as predicted in the assumption of the theory, the degree of inequality-aversion was unlikely to be negative. As mentioned in Footnote 7, five negative inequality-averse subjects were excluded from the sample observation. We examined 45 subjects that had various attitudes toward inequality-aversion. The estimation of Probit MLE(1) covered all of the observations of the 2,520 individual decisions due to the observations in the first round being excluded. Among these observations, the subjects decided to join 1,692 times, while they did not decide to join 828 times.

The inequality-averse attitudes, the dummy variable of being critical players, and the decision made in the previous round had significant positive effects on the decisions. This interesting result showed that the strongly inequality-averse subjects were more likely to participate in a coalition. It implied that subjects participated as the experimental design suggested, and that their decisions were consistent. Having said that, the group contribution to the participation was significantly negative due to the fact that the free-riding incentive was higher when the coalition was expanded. Another interesting result was that the subjects with a weaker religious belief were more likely to participate.

This experimental design used a number of critical players to form a profitable coalition. Those critical players were essential in order to stabilise the coalition internally. Probit MLE(2) and Probit MLE(3) examined the observations of the critical players. Eighty five percent out of the 1,500 observations participated in a coalition as the design suggested. Their decisions were consistent with the past decisions. On the other hand, the experimental results showed that the coalition instability was caused by the subjects with low degrees of inequality-aversion rather than those with high degrees of inequality-aversion. As discussed in the theoretical section, egalitarians might break a coalition internally. In contrast, the experimental evidence showed that stronger inequality-aversion led subjects to stabilise the coalition internally.

Interestingly, pro-left-wingers were more likely to participate. That being said, subjects had stronger incentives to form a profitable coalition when they were egalitarians or pro-left-wingers. Perhaps a low profit, but safe action appeared to be more favourable than a risky strategy of punishing and forcing free-riders to participate.

Having discussed the critical players, the non-critical players were assessed by the estimations of Probit MLE(4) and Probit MLE(5). Those non-critical subjects had the free-riding incentives. The results showed that such incentives were rejected for nearly half of the 1,200 observations. Again, the decisions were consistent with the historical data. Besides, egalitarian subjects were more likely to compromise and cooperate. Subjects with stronger attitudes towards inequality-aversion, such as taking the fair allocation for more than 6 rounds, were more likely to participate in a coalition.

Apart from the inequality-averse attitudes, the estimation of Probit MLE(5) examined the factors. In contrast to the experimental evidence of the study by Burger and Kolstad (2010), the results of this study did not support their earlier finding that higher marginal benefits would significantly increase a coalition size and, consequently, the total contribution. The results suggest that the free-riding incentive for non-critical players could be mitigated by a lower marginal benefit. In other words, the marginal benefit to the total abatement had a significantly negative effect on the willingness of participation. It was intuitive that the non-critical players would not participate when the incentive was high.

There is more policy implication from our results. Individual political and religious attitudes had significant effects on the willingness of participation. The pro-left-wingers played strategically by cooperating when they were critical and not cooperating when they were non-critical. On the other hand, the atheists were more likely to cooperate when they were non-critical.

5. Conclusions

In this study, we examined the impacts of social preferences on the individual incentives of participating in climate coalitions by using a laboratory experiment. Theoretically, when countries are self-interested, stable coalitions exist if signatories have no incentive to leave and non-signatories have no incentive to join. In particular, we focused on a unique equilibrium which players had a weakly dominant strategy on membership decisions. Any critical player was essential to a profitable coalition and could not be replaced by all non-critical. Having introduced the benchmark model, we considered the individuals' attitudes toward inequality-aversion in this study. When countries had a strong inequality-averse attitude, egalitarians could break the coalition internally or externally. Therefore, the coalition formation might enlarge, remain the same as the results of the benchmark, or become unstable.

A set of experiments was conducted in order to validate this hypothesis. The first test measured the individuals' attitudes towards inequality-aversion. The second test was a public good game, which mimicked the international environmental convention. Subjects were given different payoff tables and asked whether or not they would join a coalition. Regarding the coalition formation in the membership game, the formation was usually larger than the self-interested equilibrium size. Under such conditions, the coalition formation was still difficult to predict, and even in this study, we employed the predictions with the individuals' inequality-averse attitudes. The predicted formation was unstable and usually larger than the actual formation.

Turning back to the research question, one may suggest that the inequality-averse attitudes had significant positive impacts on the incentives of participation. In particular, when the subjects were non-critical players, the egalitarians were likely to give up the free-riding benefit by joining a coalition. This result could explain why the coalition formation was usually larger than the Nash equilibrium.

This study has multiple implications for public opinion elicitation and public policy. As mentioned earlier, the coalition formation could be influenced by individual social preferences. Our findings suggest that it is important to highlight not only the individual payoff but also the gap between players. Some significant factors could be illustrated intuitively. Most subjects were rational since they behaved consistently and pursued monetary incentives. In terms of the coalition formation, when more players were critical to the coalition and marginal benefits were higher both could lead to a higher free-riding incentive for non-critical players. Apart from that, other factors may not have been intuitive. According to their self-examination in the questionnaire, the pro-left-wingers behaved strategically. They cooperated when they were critical players and non-cooperated when they were non-critical players. On the other hand, those atheists were more likely to be cooperative when they were non-critical players.

In conclusion, individual inequality-averse attitudes could be the reason for large coalition formation. No matter whether they were critical players or not, a stronger attitude towards inequality-aversion led to more willingness to participate in a coalition. The results of this study also suggested that the individual motivation could be affected by their political and religious attitudes. The implications could advise policy makers on constructing a climate coalition for a better future.

Reference

Alló, Maria, and Maria L Loureiro. 2014. "The role of social norms on preferences towards climate change policies: A meta-analysis." *Energy Policy* no. 73:563-574.

Bahn, Olivier, Michèle Breton, Lucia Sbragia, and Georges Zaccour. 2009. "Stability of international environmental agreements: an illustration with asymmetrical countries." *International Transactions in Operational Research* no. 16 (3):307-324.

Barrett, Scott. 1994. "Self-enforcing international environmental agreements." *Oxford Economic Papers*:878-894.

Barrett, Scott. 2001. "International cooperation for sale." *European Economic Review* no. 45 (10):1835-1850.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann. 2011. "A within-subject analysis of other-regarding preferences." *Games and Economic Behavior* no. 72 (2):321-338.

Bratberg, Espen, Sigve Tjøtta, and Torgeir Øines. 2005. "Do voluntary international environmental agreements work?" *Journal of Environmental Economics and Management* no. 50 (3):583-597. doi: https://doi.org/10.1016/j.jeem.2005.03.002.

Breton, Michèle, Lucia Sbragia, and Georges Zaccour. 2010. "A dynamic model for international environmental agreements." *Environmental and Resource Economics* no. 45 (1):25-48.

Burger, Nicholas E, and Charles D Kolstad. 2010. "International Environmental Agreements: Theory Meets Experimental Evidence 1."

Carlsson, Fredrik, Dinky Daruvala, and Olof Johansson-Stenman. 2005. "Are People Inequality-Averse, or Just Risk-Averse?" *Economica* no. 72 (287):375-396.

Charness, Gary, and Matthew Rabin. 2002. "Understanding social preferences with simple tests." *The Quarterly Journal of Economics* no. 117 (3):817-869.

d'Aspremont, Claude, Alexis Jacquemin, Jean Jaskold Gabszewicz, and John A Weymark. 1983. "On the stability of collusive price leadership." *Canadian Journal of economics*:17-25.

Dannenberg, Astrid, Andreas Löschel, Gabriele Paolacci, Christiane Reif, and Alessandro Tavoni. 2015. "On the provision of public goods with probabilistic and ambiguous thresholds." *Environmental and Resource Economics* no. 61 (3):365-383.

Dannenberg, Astrid, Thomas Riechmann, Bodo Sturm, and Carsten Vogt. 2012. "Inequality aversion and the house money effect." *Experimental Economics* no. 15 (3):460-484.

Domínguez Arcos, Fernando, Xavier Labandeira Villot, and María Loureiro García. 2011. "Climate Change Policies And Social Preferences In Galicia And Spain." *Revista Galega de Economía* no. 20 (1).

Eyckmans, Johan, and Michael Finus. 2006. "New roads to international environmental agreements: the case of global warming." *Environmental Economics and Policy Studies* no. 7 (4):391-414.

Fehr, Ernst, and Klaus M Schmidt. 1999. "A theory of fairness, competition, and cooperation." *The quarterly journal of economics* no. 114 (3):817-868.

Fischbacher, U. 2007. z-Tree: Zutich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, 171-178.

Greiner, B. 2004. The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics.

Grüning, Christine, and Wolfgang Peters. 2010. "Can justice and fairness enlarge international environmental agreements?" *Games* no. 1 (2):137-158.

Hanemann, Michael, Xavier Labandeira, and Maria L Loureiro. 2011. "Climate change, energy and social preferences on policies: exploratory evidence for Spain." *Climate Research* no. 48 (2/3):343-348.

Kolstad, Charles D. 2014. International environmental agreements among heterogeneous countries with social preferences. National Bureau of Economic Research.

Kosfeld, Michael, Akira Okada, and Arno Riedl. 2006. "Institution formation in public goods games."

Kroll, Yoram, and Liema Davidovitz. 2003. "Inequality aversion versus risk aversion." *Economica* no. 70 (277):19-29.

Lange, Andreas. 2006. "The impact of equity-preferences on the stability of international environmental agreements." *Environmental and Resource Economics* no. 34 (2):247-267.

McEvoy, David M, Todd L Cherry, John K Stranlund, Jon Hovi Cherry, and David McEvoy. 2014. "International environmental agreements with endogenous minimum participation and the role of inequality." *Toward a New Climate Agreement: Conflict, Resolution and Governance, Routledge*:93-105.

Svenningsen, Lea Skræp. 2017. Distributive outcomes matter: Measuring social preferences for climate policy. University of Copenhagen, Department of Food and Resource Economics.

Svenningsen, Lea Skræp, and Bo Jellesmark Thorsen. 2017. Preferences for distributional impacts of climate policy. University of Copenhagen, Department of Food and Resource Economics.

Willinger, Marc, and Anthony Ziegelmeyer. 2001. "Strength of the social dilemma in a public goods experiment: an exploration of the error hypothesis." *Experimental Economics* no. 4 (2):131-144.

Yang, Y., S. Onderstal, and A. Schram. 2012. Inequity Aversion Revisited. Paper read at Discussion Paper.

Yang, Yang, Sander Onderstal, and Arthur Schram. 2016. "Inequity aversion revisited." *Journal of Economic Psychology* no. 54:1-16.