Essays on the Economics of International Environmental Agreements

YU-HSUAN LIN

PhD in Economics

September 2013

Abstract

This thesis consists of three essays on the economics of international environmental agreements (IEAs). The essays provide both theoretical models and experimental evidences for investigating individual incentives of participating in IEAs based on different assumptions about preferences.

Chapter 1 and chapter 2 explore the incentives of participating in IEAs with social preferences (also known as other-regarding preferences) in a static model through experimental methods.

Chapter 1 examines the effect of inequality-aversion. The theoretical prediction for the proposed experiment in this chapter expects that the players with a high degree of inequality-averse preference will violate the internal constraint and be absent in a membership game. As a consequence, the coalition formation will become unstable. The experimental outcome confirms that a stable coalition is indeed very rare. This is because the individual preferences on inequality-aversion play a role in shaping coalition formation. However, interestingly, highly inequality-averse subjects, while following their best strategies to participate, are less likely to be absent from the coalition. According to this study, the internal constraint is mostly broken by lowly inequality-averse

subjects.

Chapter 2 investigates the effect of altruistic preferences. The theoretical prediction of this experiment is that the subjects with a high level of altruism are more likely to resist the temptation of free-riding and thereby are more likely to participate in a coalition. The experimental evidence confirms that the coalition formation is affected by individual altruistic preferences. However, the incentive of participation seems to be negatively correlated with the altruistic attitude: the lower the degree of the altruistic preference is, the more likely the subjects would participate.

Chapter 3 examines the impact of sustainability, which are considered as cross-generational social preferences, on the coalition formation in a two-stage game in two periods. This study confirms the importance of the awareness of sustainability to international environmental conventions. When the intergenerational fairness and altruism are taken into account, a coalition formation will be expanded. The numerical example indicates that the marginal cost of the total emissions is an important factor for the formation of IEAs. In contrast, the advanced level of technology development may lead a more efficient production per unit of emissions, but it also encourages countries to emit more in total and have a lower level of welfare. Only when the preference weighting attached by one generation to the welfare of the next generation is considered in international environmental conventions, a sustainable system could be succeed.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my appreciation to the supervision of Dr. Bipasa Datta, who encouraged me to complete my research and shared her life experience as a mentor. I would also like to give my heartfelt appreciation to Professor John Hey, who kindly advised me to develop my knowledge on experimental economics and provided immensely helpful comments on the draft of my thesis. Without his help and inspirations, I could hardly complete this thesis.

Another member of my thesis advisor group, Professor Peter Jeremy Simmons has also given me sound feedback. My external examiner, Professor Alistair Ulph, and my internal examiner, Mr John Bone, both made the viva an enjoyable yet unforgettable experience for me, and provided constructive feedback for me to work on. Without their brilliant comments and suggestions, I would not have brought my thesis to this high standard and equipped with fresh ideas to extend my research.

A special "thank you" to the lovely and helpful colleagues Dr. Ian Corrick, Dr. Saruta Benjanuvatra, Dr. Ryota Nakamura, and Dr. Vivien E. Burrows for supporting me in writing and encouraging me to work toward my goal.

Lastly, no words cannot express my gratitude to my beloved parents, my sister and brother for their endless love and support.

Author's declaration

All three chapters in this thesis are single-authored. The experiment in chapters 1 and 2 was conducted at the centre for EXperimental EConomics (EXEC) laboratory at the University of York in 2013. It was a joint project with Dr. Bipasa Datta and financially supported by the Research Priming Fund at the University of York. The initial result was presented at the North American Economic Science Association Conference at Arizona in November 2012, at the 2nd White Rose Doctoral Training Centre Economics Conference at York in February 2013, and at the 5th World Congress of Environmental and Resource Economics at Istanbul in July 2014 .The early version of Chapter 3 was presented at Belpasso International Summer School on Environmental and Resource Economics at Belpasso in September 2012, oikos Young Scholars International Economics Academy at Geneva in September 2011, and at the Korea and the World Economy XIII conference at Seoul in June 2014.

I deeply appreciate to the comments from Professor Charles D. Kolstad, Professor Kiyoun Sohn and many participants in the conferences. Their valuable comments enhance this thesis. However, all the errors and mistakes belong to me.

(First version: September 2013

Revised version: January 2015)

Introduction

It is widely recognized that the ecosystem on the earth has changed dramatically over the last few decades due to rapid economic and industrial development. Crafting solutions to balancing development and sustainability has been urged by many international organisations, such as the United Nations (UN), the Organisation for Economic Co-operation and Development (OECD), and the World Bank. It was thought that some environmental issues could be dealt with locally, for example, at the national level. However, some environmental issues (such as water and air pollution, generation of solid and hazardous waste, soil degradation, deforestation, climate change and loss of biodiversity) usually are so complex and so widespread (across sovereign borders) that they require collaboration between states. In this respect, inter-governmental law-making and multi-disciplinary international research are vital and have become common practices to tackle these complicated socio-environmental problems. International environmental agreements (IEAs) are one of the mechanisms constructed to regulate and manage the situation. According to Mitchell (2003), there are over 700 multilateral agreements and over 1,000 bilateral agreements. IEAs have become the most important mechanism for solving the international

environmental problems.

IEAs and international conventions have endeavoured to deal with a wide range of environmental issues such as climate change, biological diversity, control of movements of hazardous wastes, and ozone layer. These topics can be categorised into two main types: natural resource sharing and reducing international environmental damage. The former targets at solving local environmental issues related to limited natural resources (e.g., water, fisheries, timber and other elements of the natural world), while the latter deals with global environmental concerns (e.g., acid rain, sea pollution, ozone layer depletion, climate change, global warming).

IEAs concerning natural resources are often bound by geographic boundaries (e.g. the atmosphere, rivers, lakes, oceans, and terrestrial habitats), thereby the discussion usually stays at a regional level. The well-known subcategory objectives related to these IEAs include freshwater resources (Convention on the Protection and Use of Transboundary Watercourses and International Lakes, 1992), marine living resources (Common Fisheries Policy of the European Union, 1970), terrestrial living resources (Convention on Biological Diversity, 1992; International Tropical Timber Agreement, 1994), and marine environment resources (United Nations Convention on the Law of the Sea, 1982; International Convention for the Prevention of Pollution From Ships, 1983).

These resources, such as fish stocks, timber and coal, are common goods. Being rival goods means that the proposals need to consider how to best utilise these non-excludable goods, and identify an optimal amount of consumption. If the resources are renewable, such as forests and fisheries, then the proposals would centre on how to sustainable harvest the resources by making sure that

the quantity of consumption does not exceed the rate of regeneration. If the resources are non-renewable, such as oil and gas (i.e., the stock of the resources is finite), then the proposals would focus on identifying an optimal use that aims to extend the length of exploitation. The mechanisms usually are engineered in a way through which most common goods can be transformed into some forms of revenues (e.g., fiscal, economic or other values) so that membership countries are motivated to participate in this sort of IEAs.

A well known category regarding international environmental problems less bound by national borders is the IEAs that aim to reduce environmental damages by developing a cleaning-up mechanism and an emissions abatement scheme. The subcategory objectives related to these IEAs include ozone layer depletion (Montreal Protocol, 1997), climate change (the United Nations Framework Convention on Climate Change (UNFCCC), 1992), and acid deposition (Protocol on the Reduction of Sulphur Emissions, 1985). The objectives of these IEAs are to manage public goods, which are non-rival and non-excludable. Because of the characteristics of non-rival goods, free-riding is unavoidable and is a key issue in this sort of IEAs. Negotiations are more challenging than those in the first category dealing with natural resources.

This thesis contributes to the discussion on motivations for countries to participate in IEAs. More precisely, the focus is placed on the agreements of abatement of carbon and greenhouse gas emissions. A substantial amount of literature has approached this topic by examining the effect of policy instruments, such as punishment scheme, sanctions and side payment, and how they enhance a stable coalition. However, the existing discussion mostly centres on the design of IEAs frameworks rather than on the motivations of forming and participating in IEAs. In order to rethink IEAs and the formation processes,

the chapters in this thesis explain what motivates countries to take part in an IEA without any policy instruments.

Literature review on the studies of international environmental agreements

Since the environmental issues are at the top of the global policy agenda, there has been an increasing amount of literature on international environmental agreements. In terms of the methods, the studies can be roughly classified in three main fields: game theory, calibration, and experimental method.

Game theory has been a common methodology for analysing the formation and stability of IEAs. If we consider international environmental issues as public goods, there exist two main problems: free-riding and externalities. No individual can be excluded from the other's transboundary environmental damage, nor can anyone share the benefit of pollution abatement. Finus (2008) claims that game theory is the ideal tool to study IEAs because 'game theory is a mathematical method that studies the interaction between agents based on behavioural assumptions about the preference of agents and makes prediction about the outcome of these interactions by applying various equilibrium concepts (Finus, 2008)'.

Due to the limitations of data collection, very few empirical studies examine policy effects with empirical data. For instant, Bratberg *et al.* (2005) employ the double difference method to examine the policy efficiency of Sofia Protocol during the period 1985-1996. Their empirical evidence shows that the estimated yearly reduction in nitrogen oxides is nearly 2.1% higher than it

would have been without the Protocol. Nevertheless, knowledge about global environmental issues, such as climate change, is still very limited.

In order to estimate the impacts of the climate change policies, DICE (Dynamic Integrated model of Climate and the Economy), developed by Nordhaus in 1994, was the first integrated-assessment model and the most widely used in the economics of climate change. This model evaluates different climate change strategies by its general equilibrium approach. It has inspired the development of several recent models, including MERGE (Model for Evaluating Regional and Global Effects of GHG reductions) and STACO (STAbility of COalitions) model. The MERGE model is developed by Manne *et al.* (1995). It is a fully integrated applied general equilibrium model with a flexible design to evaluate the impact of climate policies on a wide range of contentious issues, e.g., costs and benefits of mitigation policies, valuation and discounting issues. STACO is a game-theory-based project on the formation and stability of international climate agreements. The initial model was built by Finus *et al.* in 2006. A game-theoretic framework has provided this model the ability to analyse not only the interactions between players but also the stability of potential international climate agreements. Several topics have been investigated, e.g., stability of climate coalitions in a cartel formation game, exclusive membership, multiple coalition games, transfer schemes, quota, the stability likelihood of coalitions under uncertainty, technological change, and sequential games.

These calibrations are powerful research tools for estimating the influences of climate policies. However, they are constrained by model assumption and exogenous scientific parameters. Recently, experimental research has proven useful in evaluating policy instruments, particularly when empirical data is

prohibitively costly (Eckel and Lutz, 2003). A growing number of literature in experimental studies indicates that appropriately designed and tested policy mechanism may help to alleviate environmental problems and provide useful advice to policy makers (Bohm, 2003; Pevnitskaya and Ryvkin, 2011).

Barrett (1994) provides a seminal study that positions 'self-enforcing' as a key incentive for participating and interacting in IEAs. His key assumption of the absence of a supra-national body to structure an IEA leads him to suggest that participation is voluntary and all countries are free to enter or to withdraw from a coalition. While an IEA aims to maximise the aggregate net benefit, individual nonsignatories aim to maximise their own net benefit. In joining an IEA, signatories receive a reward from acceding to the agreement and avoid the punishment from withdrawing. Non-signatories may be penalised but also enjoy the free-riding benefit. The majority of the literature, however, follows D'Aspremont *et al.* (1983) who argue that a stable coalition has two constraints: the internal one where no signatory has any incentive to withdraw from the coalition; and the external one where no nonsignatory has any incentive to join the coalition.

Cross-border or macro-regional environmental issues concern public goods, in general. One of the key characteristics of public goods is the free-riding effect. When a profitable coalition is formed, all other countries outside the coalition would receive positive externality from this grouping. In order to minimise this effect and maintain a stable self-enforcing coalition, an efficient policy mechanism is desirable.

To find a well-designed IEA, several policy mechanisms have been discussed in the literature and launched in practice. First of all, punishment schemes are widely applied to existing IEAs. The majority of the literature in both theo-

retical and empirical studies considers punishment schemes through economic modelling to understand IEAs membership (e.g. Bahn *et al.*, 2009; Barrett, 1994, 2001; Breton *et al.*, 2010; Lessmann *et al.*, 2009). Both theoretical and empirical studies show that the absence of punishment schemes results in a significant disincentive to be a signatory of an IEA.

Secondly, the side payment mechanisms have had great influences on the structure of IEAs. Given a group of countries which have been committed to cooperate, it is in principle possible to achieve Pareto improvement with a side payment mechanism to encourage the nonsignatories to reduce their emissions in exchange for transfers from the signatories. In other words, the mechanism allows side payments from coalition members to non-members. However, Hoel and Schneider (1997) argue that the proposal of offering disengaged countries a transfer to reduce their emissions (provided that the country does not commit itself to cooperation) tends to reduce the incentive of the receiving country to commit itself to cooperation. They emphasise that the side payment is a disincentive for participation in an IEA. Also, total emissions will be even higher in situation where side payments are in place to allow transfer and offset than those without.

Different transfer mechanisms would lead to various results. In practice, the transfer mechanism known as the "joint implementation" in the Kyoto Protocol allows a country with an emissions reduction or limitation commitment to earn emissions reduction units from another signatory. This transfer mechanism allows the transfers of emissions permit among coalition members. Some calibration studies of Nagashima *et al.* (2009) and Dellink and Finus (2012) appraise such transfer mechanism can stabilise larger coalitions and increase global abatement levels.

Thirdly, the Emission Trading Scheme (ETS), a well-known market-based approach, has been launched in 2005 [1]. The ETS mechanism allows countries to trade six major greenhouse gases permit among signatories. These trades allow transfers among countries in the coalition. The transfers imply that if a signatory reduces its emissions more than the required amount for achieving the assigned emissions permit level, the country can sell permits to other signatories. McKibbin *et al.* (1999) examine the effects of the tradable emissions permit system proposed in the Kyoto Protocol with an estimated multi-region, multi-sector general equilibrium model of the world economy. Their results suggest that capital flows significantly affect the domestic effects of the emissions mitigation policy. However, Karp and Zhao (2010) claim that the policy effect of the ETS is ambiguous. Only with an escape clause policy and a safety valve policy could the ETS have a significant effect on enlarging the equilibrium level of abatement and the number of signatories.

Whilst this thesis acknowledges the importance of the transfer mechanisms that is not of the scope of this thesis to explore policy mechanisms. For the purpose of studying motivations and behaviours in the membership game, this study simplifies the allowance of the emissions permit to be non-transferable. However, our design requires wealth transfers among member of the coalition. In other words, signatories with high marginal benefit of the total emissions have to financially assist those signatories with low marginal benefit. This strong assumption implies that coalition members share equal responsibility to maximise the collective payoff. The detail will be discussed in the section of model setting in Chapter 1.

[1] European Union ETS was the first large emissions trading scheme in the world. It was launched in 2005 to combat climate change and is a major pillar of EU climate policy.

Although Barrett (1994) provides a fundamental explanation for the difficulty in making a sustainable and stable IEA, there exist three major assumptions that limit his arguments.

The first assumption, which states that 'all countries are identical', does not correspond to the reality. This assumption suggests that the formation of IEAs depends on the marginal benefit of total abatements. This assumption leads to some oversimplified results when describing some given scenarios as will be seen in Chapter 3. These results are so oversimplified that likely to be disconnected from the reality. The assumption of the participation of heterogeneous countries has received more attention in recent studies. This assumption of heterogeneous countries participating in IEAs helps underline the asymmetries, the reactions and behaviours of countries with diverse interests and characteristics.

But there are different ways of categorising heterogeneous countries. For example, Barrett (2001) categorises asymmetric countries into 'rich' countries with more ozone-depleting substances and 'poor' countries with less substances. He suggests that the rich countries can contribute more to the environment with their greater ability to pay and/or their larger influence on global emissions abatement. The poor have neither the ability to pay nor the global influence. They may be suffering from immediate and severe effects of the environmental damage, while the rich face a smaller level of damage. Barrett's results show that stronger asymmetry between players would strengthen the willingness to participate in an IEA. His finding is supported by Dellink and Finus (2012) who conduct a study on transfers of emission permits within IEAs.

There also exists a huge volume of literature, including Bahn *et al.* (2009),

which discusses countries with various marginal environmental damage costs. Heterogeneity is also discussed in empirical studies, e.g., the experiment conducted by Burger and Kolstad (2010) that examines the theoretical works with two-type marginal benefit of total contribution. Although the design with two-type marginal benefit could distinguish different forms of participation, the key question regarding individual incentives to participate has not yet been answered. In other words, their studies did not illustrate why subjects with the same type of marginal benefit make different decisions. In order to observe individual decisions, Chapters 1 and 2 are based on an experiment built on an environment in which diverse marginal benefits are considered.

The second weakness in Barrett (1994)'s paper is the assumption of perfect information. This assumption ignores factors that may lead to imperfect information and thereby is incapable of capturing uncertainty. Accurate information is necessary for making international, especially global environmental policies. Given their complexities, environmental problems are hardly well explained by the most advanced science, let alone well-known to decision makers who are involved in the negotiation of abatements. For example, contradictory scientific evidences and arguments for climate change have been observed over the last decades (e.g. House of Lords, 2005). The disparity of evidences and debates lead to the ambiguity of preferences of the general public.

Since the scientific evidence on the impact on the ecosystem is ambiguous, a perfect far-sighted decision-making process does not exist. Not only are limited information and uncertainty crucial, so is how these factors shape decision makers' strategies and behaviours. In order to study the implications of uncertainty, previous studies have fixed the distribution of the random parameters and the operational patterns to specify how agents form their expectations

(Finus and Pintassilgo, 2013). In addition to that, strategic decision makers adapt to cope with uncertainty and collect more information to facilitate their decision making. In order to model the learning process and its effect, Finus and Pintassilgo (*ibid*) also take 'time' into account. They argue that timing is important in the learning process. Learning takes place when the information about probability of heterogeneity is revealed either *ex ante* or *ex post* to countries. No learning takes place when countries know the information after making decisions; complete learning takes place when they know the information before making decisions. If environmental threats are not as serious as scientists predict, some over prepared solutions will lead to unnecessary waste. On the other hand, if threats are more severe than expected, more actions have to be taken to cover the loss for not enough preparation. This possible loss in the future is far larger than the spending on the protection in the present. Besides, the key point is that most environmental damage is irreversible, such as ozone layer depletion. Bearing such irreversibility in mind, a decision maker who has no information may prefer over-protection to no preparation.

Kolstad (2007) and Kolstad and Ulph (2008) consider the effects of the learning process and irreversibility in a single decision maker model conditioned by uncertainty. They assume that players could have two types of learning processes: partial learning and complete learning. They argue that uncertainty in a complete learning process leads to more cooperation but lower aggregate net benefits than in an environment where no learning takes place. Partial learning would lead to lower membership and even lower expected aggregate net benefit. Their findings, surprisingly, show that certain information has a negative effect on IEAs. Helm (1998) provides an explanation for this: countries can use the veil of uncertainty to hide their distributional interests

and lead to the success of IEAs without engaging in any learning process. Dellink and Finus (2012) investigate uncertainty with their simulation on climate change. They find that learning processes (both complete and partial) can only be positive if emission permit transfers are considered.

To specify and enhance our research questions, the design of the public goods game in our experiment provides information with regard to the subjects' own payoffs as well as those of others. But even so, the outcomes may not be as consistent as the Nash predictions.

The last fundamental but questionable assumption in Barrett's model is that agents are egoists. In light of the Nash equilibrium, this implies that a rational agent would choose the highest payoff. The assumption has been widely employed in the majority of the theoretical studies of IEAs (e.g. Barrett, 2001; and Breton *et al.* 2010). However, recent experimental evidences have suggested that the assumption of egoistic preferences is not enough to explain individual decision makers' behaviours in an interactive game (Kosfeld *et al.*, 2009; Burger and Kolstad, 2010). These studies claim that people are far less likely to free ride and more likely to cooperate than the egoistic prediction assumes. Hence, social preference (or other-regarding preference) has been proposed in recent studies (e.g. Kolstad, 2014) to address this gap. This study follows this trend of thought and considers two types of other-regarding preference, namely inequality-aversion and altruism, to develop the model and experimental design.

Structure of the Thesis

This thesis consists of three original studies on the economics of international environmental agreements (IEAs). The focus lies in individual behaviours and decision-makings of IEAs. Three hypotheses are proposed and tested. All studies intend to contribute to theoretical as well as current policy-related discussion. In terms of the latter, when being applied to real-world policy-making, it is anticipated that a better understanding of will help tackle environmental issues efficiently, thereby reduce the risks of environmental disasters and enhance human welfare.

The three hypotheses to be tested in the thesis can be divided into two themes from a methodological perspective.

The first theme, featuring Chapter 1 and Chapter 2, employs experimental methods to examine the effectiveness of the theoretical prediction of other-regarding preferences on the static formation of IEAs. Individual payoffs in a membership game are deemed mutually affected. Chapter 1 focuses on the effect of fairness, which shapes the payoff gaps between agents. Chapter 1 investigates agents' decision-making in a membership game which are not as Nash equilibrium predicts when heterogeneous preferences on inequality-aversion are presented. In our theoretical model, in order to achieve fairness, agents who have a higher degree of inequality-aversion are more likely to punish free riders by leaving a coalition. Unlike what has been suggested in the existing literature, the prediction on the formation of an IEA could be equal to or larger than the Nash prediction, or be an unstable coalition.

To explore the effect of inequality-averse preferences on cooperation, an experiment with two stages is conducted. Before playing first of the experiment,

which imitates an IEA formation, subjects are asked to take an inequality-averse test which indicates their individual social preferences. Chapter 1 assumes that subjects care about not only their own payoffs but also the gap between their own payoffs and those of others. In other words, subjects consider the variances of individual payoffs.

Chapter 2 investigates how the levels of altruism shape the incentives to free ride. We assume that agents may have different altruistic preferences which influence their decisions in a coalition game. Similar to the design in Chapter 1, an individual altruistic test is provided before a public goods game. Subjects are expected to consider not only their own payoffs but also the overall payoff of all subjects.

The result shows that, in order to enlarge the overall welfare, agents have strong altruistic preferences would give up the free-ride rewards. Our theoretical prediction claims that the formation of an IEA could be equal to or larger than the Nash prediction.

The experimental evidences in Chapter 1 and Chapter 2 make two key contributions to the existing discussion. Firstly, they provide a novel exploration of individual behaviours in a IEA, based on a case study of unique equilibrium coalitions. Although it is difficult to generalise from a case study, it still helps to identify behaviour patterns of individual decision-makers since each subject has a weakly dominant strategy to determine their status in a membership game. Secondly, the experiments examine and verify the theoretical predictions with two types of other-regarding preferences.

Both the experimental evidences and theoretical predictions confirm that the willingness to participate in IEAs is significantly associated with the degree of inequality-aversion and the degree of altruism. However, the experimental

results are against the hypotheses: the results in Chapter 1 show that the lower the degree of inequality-aversion a subject has, the more likely the subject is to behave strategicly. When subjects' dominant strategies are to join the coalition, those having a lower degree of inequality-averse preference are more likely to punish free-riders by leaving the coalition. When their dominant strategies are not to join, the subjects with lower degree inequality-aversion have higher willingness to cooperate. The results in Chapter 2 illustrate such strategic behaviours in the membership game. Subjects who have lower degree of altruistic preferences are more likely to cooperate in the public goods game. Overall, the experimental evidences show that subjects' decisions differ and change because of their social preferences.

Chapter 3 is a purely theoretical study. Unlike the static decision discussed in the first two chapters, Chapter 3 aims to explore the causal relationship between the preference weighting to the welfare of the next generation and the incentives of participating in IEAs. In order to examine the cross-generational effect, this chapter creates a two-generation model which describes the decisions made by the present generation who may or may not take the welfare of the future generation into account. In this model, sustainability is defined by the criterion that the welfare of the future generation is not worse than that of the present generation. The study aims to find the emissions level and coalition formation in different policy contexts.

By evaluating the impacts of the cross-generational fairness and altruism on the formation of IEAs, Chapter 3 identifies the importance of the perceptions of '*sustainability*' to IEAs. We substantiate the concept of 'sustainability', a common (and perhaps over-loaded) buzzword often used at international environmental conventions. In so doing, we provide economic explanations

about how cross-generational payoffs can be maximised and how to extend the length of consumption of limited resources. Unlike international financial and monetary agreements which focus on on-going real-world conflicts and issues, IEAs are created to avoid possible disasters in the future which are difficult to predict. This chapter offers an economic explanation for some characteristics of IEAs.

The numerical examples in Chapter 3 show that when the future generation is concerned by the current one, a country is more willing to participate in an IEA. However, the discount factor attached by one generation to the welfare of the next has small and ambiguous impact on the coalition formation. The technology level has a positive effect on the emission level, but not on the formation size. In other words, the level of technology level may not be the key factor that mitigates the free-riding effect, because an efficient technology could also increase the incentive of emitting. On the other hand, the marginal cost of total abatement has negative impact on the emissions level. A grand coalition is possibly formed when the marginal cost is very small.

The thesis concludes with discussions on future studies extended from the lessons learned from investigating individual behaviours of dealing with IEAs memberships.

Chapter 1

Inequality-Averse Preference for International Environmental Agreements

1.1 Introduction

International environmental agreements (IEAs) are typically viewed as coalitions of agents providing public goods (e.g., abatements of greenhouse gas emissions). Since the publication of Barrett (1994), the literature on IEAs by and large assumes that countries self-enforce themselves to join an IEA. It means that countries sign an IEA for economic reasons. A stable IEA exists under both internal and external constraints. When the payoff of being a signatory is better than that of being a nonsignatory, a country has an incentive to participate and the coalition is stable internally. On the other hand, when a nonsignatory has no incentive to join the coalition and decides not to participate, the IEA is stable externally. From the macroscopic perspective, a

robust IEA requires both internal and external stability. It is a state where no insider wants to leave and no outsider wants to enter. Nevertheless, the incentives of an individual agent have not been fully examined from the microscopic perspective in the existing literature. Although the majority of the studies on IEAs has investigated incentives (e.g. Barrett, 2001; Finus, 2008), their main focus is on the formation of IEAs from a *macroscopic* perspective. Individual incentives have been over-simplified in the literature. There may exist several equilibria, individual incentives are not clear even when a coalition is stably formed.

This chapter discusses individual incentives of joining a coalition, and their roles in an interactive game. The interaction between agents is closely linked with agents' individual preferences. This study employs the *microscopic* perspective to explore how individual preferences shape decision making.

In the existing literature, two issues still await to be addressed: the arguably unavoidable free-riding effect and a presumed egoistic preference.

Free-riding has largely been considered as the most important obstacle for the formation and existence of successful IEAs. This is the main reason why, a large IEA is not easy to be formed without any policing mechanism, in light of the Nash equilibria static game. However, recent experimental evidences on IEAs suggest that people are far less likely to free ride and more likely to cooperate than the theory suggests (Kosfeld *et al.*, 2009; Burger and Kolstad, 2010). But why this is so has not been well-explained by the models in the literature.

Furthermore, existing research findings on IEAs largely presume that an individual's preference is egoistic/selfish. However, the solutions to international environmental problems require cooperation and interaction between

different nations at a global scale so as to prevent environmental or natural disasters or damages from happening. International cooperations are called for to deal with global issues. In such interactive game with common goal to minimise the loss of the society and environment, the assumption of a pure egoistic preference may not be enough to capture players behaviours.

Some have suggested to address this limitation by taking the role of other-regarding preferences (also known as social preferences) into account. Kosfeld *et al.* (*ibid*) employ the inequality-averse preference (proposed by Fehr and Schmidt, 1999) and confirm with laboratory-based evidence that when inequality-averse players exist, the coalition is no longer a Nash prediction, and the grand coalition becomes an expected equilibrium outcome. On the other hand, Kolstad (2014) adopts Charness and Rabin's (2002) social preferences theory which suggest that agents mainly care about three things : private payoff, fairness in payoffs, and overall efficiency. In contrast to the finding of Kosfeld *et al.* (*ibid*), Kolstad argues that the size of an equilibrium of a coalition is smaller when social preferences exist.

Although the coalition formation with social preference has been examined in the literature, its influence on individual behaviours in an interactive coalition has not been fully explored. In other words, individual incentives for participating in a coalition are still unclear. This is partly due to the fact that economic models usually are based on several assumptions to reduce uncertainties and ambiguities. But these assumptions make capturing individual incentives difficult. For example, even with the assumption of heterogeneous agents, players were given the same payoff table in an experiment. There exist multiple equilibria and several possible coalition combinations, individual incentives are not possible to be predicted.

To address these gaps in experimental studies, eight particular treatments which have unique equilibrium coalition are employed in this study. In these treatments, each agent has a weakly dominant strategy to follow. The individual preference is therefore identifiable and can be observed.

This design offers two main advantages: firstly, this study endeavours to investigate incentives for participating in IEAs. If a coalition has more than one equilibrium, individual decisions cannot be predicted. But, if we have a coalition with a unique equilibrium, it would provide a suitable environment to observe individual decisions when every player has a best strategy to make. Secondly, the hypothesis of this study assumes that the other-regarding preference would influence the equilibrium differently from the egoistic preference. This entails that a coalition would be formed differently when individuals care about others agents' payoffs.

To the best of our knowledge, what motivates individuals to participate in a public goods coalition has not yet been fully explored in the existing literature. This study asks the following questions: Does the concern about fairness change players' decisions? If so, how much would they care? How do individuals' social preferences affect their own incentives for participating in a public good game?

To answer these questions, we have designed an experiment as follows. It comprises of two parts: the first part aims to find out the individual inequality-averse preference. The subjects of the experiment are paired and asked to choose from a certain fair payoff and an all-or-nothing payoff. When the expected payoff is higher than the fair payoff, those who prefer to have the fair payoff would be considered as inequality-averse players. They would be more likely to break the internal and external constraints in the coalition game.

The second part is a public good game. The subjects are grouped into 5-player groups. Since our main interest lies in the formation of IEAs, the experiment has taken out the abatement game, and turned it into a public good game which mimics the membership decision process. The subjects are given particular payoff tables to decide whether or not to join the coalition. Bearing in mind the results from the first experiment, the predictions with the other-regarding preferences are expected to explain a smaller free-riding effect and various coalition combinations.

Our theoretical finding predicts that, if the internal and external constraints hold and the condition for the unique equilibrium is satisfied, the coalition formation could be either a unique n^*-member coalition, or a unique coalition which is larger than n^*, or an unstable coalition with different inequality-averse preferences. The constraints could be violated when agents have strong attitude of inequality-aversion. However, our experimental evidence does not fully support the theory. In terms of the individual decisions, when subjects could free-ride, those with a higher marginal benefit were less likely to join a coalition and prefer to have a lower payoff. On the other hand, the subjects with a high degree of religious belief were more likely to be free-riders by not joining a coalition and having higher payoff.

From the questionnaire in the experiment, we learnt that right-wingers are more likely to build a larger coalition when they could be free-riders. Comparing to the results on the internal constraint, right-wingers are more likely to violate both internal and external constraints. Right-wingers tend to act strategicly by punishing and compromising when they are in different roles.

The chapter is structured as follows. After the introduction, in Section 2, we will compare a benchmark model based on the assumption of homogeneous

players with the model we develop based on heterogeneous players and a unique equilibrium coalition. In Section 3, the data from two experiments which are based on the theory discussed in Section 2 will be presented. In Section 4, we discuss the implications of the model and possible applications, and conclude. The theoretical proofs, the instructions of the experiment are included in the appendix.

1.2 The model

1.2.1 Benchmark model with heterogeneous players

Supposed there are N countries with different marginal benefit of total abatement, we label them as country 1, 2, ..., N. There are now $2^N - (N+1)$ possible coalition combinations [1]. In order to clarify, we assume that player 1, 2,..., n are in the group to form an IEA, player $n+1$, $n+2$, ..., N are not [2]. We rank n countries in the coalition according to the value of their marginal benefit of abatement going from high to low as $\gamma_1 > ... > \gamma_n$. On the other hand, the nonsignatories are also ranked from high to low as $\gamma_{n+1} > ... > \gamma_N$ Any marginal benefit of total abatement (γ_k, $\forall k \in [1, ..., N]$) is in the range between 0 and 1[3]. The unit cost of abatement for each country is assumed as 1.

[1] Any coalition needs at least 2 players. No coalition is a possible solution if no one cooperates.

[2] Any coalition needs at least 2 members, so $n \in [2, N]$.

[3] The meaningful range of the marginal benefit of total abatement (γ_k) is between 0 and 1. When γ_k is too large ($1 \leq \gamma_k$), an IEA is unnecessary because players already have the incentive to abate fully. When the aggregate marginal benefit is too small ($\sum_{k=1}^{N} \gamma_k \leq 1$), a profitable IEA is also non-existent because all players would pollute anyway. When the marginal benefit is in between, there may exist stable coalitions where signatories abate and nonsignatories pollute.

Each country faces a game which run in two stages: at the first stage, players play a membership game where they decide whether to participate in the coalition or not. At the second stage, given the decision made at the first stage, signatories and nonsignatories play the emissions abatement game respectively. Each nonsignatory makes her own decision on emissions abatement with the objective of maximising her individual payoff. Meanwhile, members follow a common decision on abatement with the common objective of maximising the coalition payoff. We solve this two-stage game by backward induction.

We start with the abatement game. Let any nonsignatory j's abatement be denoted by x_j. In order to simplify the model, the cost and benefit functions are both linear and the normalised level of abatement (x_j) is in the range between 0 (implies full pollute) and 1 (implies full abate).

With a profitable n-member coalition, a nonsignatory j's payoff π_j is maximised by choosing its abatement level (x_j). The problem of the nonsignatory j is as follows:

$$\begin{aligned} \max_{x_j} \pi_j &= (-x_j) + \gamma_j X \quad \forall \text{ nonsignatory } j = n+1, ..., N \qquad (1.1) \\ \text{where } X &= \sum_{i=1}^{n} x_s + \sum_{j=n+1}^{N} x_j \end{aligned}$$

where x_j is the individual abatement with its marginal benefit rate γ_j [4]. X is the total abatement which includes n signatories' aggregate reduction ($\sum_{i=1}^{n} x_s$) [5] and $(N-n)$ nonsignatories' aggregate reduction ($\sum_{j=n+1}^{N} x_j$). From the first order condition of (1.1) with respect to x_j, the optimal abatement level for a

[4] $\gamma_j \in \{\gamma_{n+1}, ..., \gamma_N\}$

[5] Because members in the coalition move as one, the aggregate emission abatement would be $\sum_{i=1}^{n} x_s = n \cdot x_s$.

nonsignatory j is doing no abatement ($x_j = 0$).

For any signatory i, all members act as one to maximise the coalition payoff and share this coalition payoff equally. The n-member coalition payoff (Π^s) is the overall pre-redistribution payoff of all members (π_i, $\forall i = 1, ..., n$). The coalition payoff is maximised by choosing the common abatement (x_s). The problem of the coalition is as follows:

$$\begin{aligned} \max_{x_s} \Pi^s &= \sum_i \pi_i \\ &= \sum_i^n [(-x_s) + \gamma_i X] \end{aligned} \tag{1.2}$$

From the first order condition of (1.2) with respect to x_s, we have

$$\frac{\partial \Pi^s}{\partial x_s} = -n + n \sum_i^n \gamma_i = 0 \tag{1.3}$$

When $\sum_i^n \gamma_i < 1$, polluting is the best strategy but then the coalition would be meaningless. To form a profitable coalition, the total contribution should go beyond the threshold which the sum of marginal benefit of members is lager than 1 ($\sum_i^n \gamma_i \geq 1$) and the best strategy for all members is fully abating ($x_s = 1$).

Since the coalition aims to maximise its payoff, individual decisions of members should achieve this goal. Burger and Kolstad (2010) note that majority voting rule, unanimity and joint payoff maximisation are all equivalent under the assumption of homogeneous agents. However, with heterogeneous agents, they suggest that majority voting reflects the interests of the median voter and may not reach a joint payoff maximum. Although wealth transfers among

member of the coalition is often suggested as being politically infeasible, Kolstad (2014) states that "sharing the wealth" within the coalition might be appropriate.

Hence, to achieve the goal of maximum a coalition payoff, each member would share the same responsibility. We assume that the coalition payoff is equally shared by all signatories. Any signatory i with a n-member coalition has a post-redistribution payoff

$$\pi_s = \frac{1}{n}\Pi^s \tag{1.4}$$

It should be noted that a rule of the coalition requires coalition members using transfers to equalise net payoffs between agents. Such rule achieves a less unequal distribution of payoffs through transferring. This assumption implies that for the main purpose of this chapter, it is difficult to separate out the issue of IEA formation and its impact on fairness from the fact that the IEA is itself a mechanism for achieving a less unequal distribution of payoffs through using transfers. Countries with higher marginal benefit of the total abatement are more likely to leave the coalition *ex post*, because those countries could earn higher payoff for the absence. However, we assume that countries have the full information when they agree to participate in an IEA, they know the consequence of being signatories and nonsignatories. Signatories will commit to stay in the coalition and make transfer to equalise individual payoffs. We appreciate that this is a strong assumption[6]. However, considering each mem-

[6]The rule would deter a country to abandon its commitment on membership by some policies, e.g. high penalty punishment and international sanction.

The issue of different policy instruments of transfer and commitment could be discussed by further studies.

ber have to move as one to maximise the coalition payoff, every member would share equally responsibility. Hence, our design of sharing the coalition is still an adequate solution.

Hence, the post-redistribution payoff of a signatory i in a profitable coalition is

$$\pi_s = -1 + \sum_{i}^{n} \gamma_i \tag{1.5}$$

In the membership game, players are asked to decide to participate in a coalition or not. The decisions are made simultaneously. With the internal and the external constraints by D'Aspremont *et al.* (1983),

$$\text{Internal constraint} \quad : \quad \pi_n^s(n^*) > \pi_n^{ns}(n^* - 1) \tag{1.6}$$

$$\text{External constraint} \quad : \quad \pi_N^s(n^* + 1) < \pi_N^{ns}(n^*) \tag{1.7}$$

There exist stable coalitions. The *internal constraint* (1.6) denotes that a signatory has no incentive to leave the n^*-member coalition and n^* is the stable number to maintain the coalition. If it is satisfied, every one would like to participate in the coalition. The *external constraint* (1.7) describes that a nonsignatory has no incentives to participate in a coalition as the $(n^* + 1)$-th member. If it is satisfied, all nonsignatories do not want to participate [7].

[7] The stability of the coalition can be explained with two 3-player cases. In case (i), if the aggregate marginal benefit of total abatement is too small to form a profitable coalition, there is no stable IEA. For example, when the set of the marginal benefit of players 1, 2 and 3 is $\{0.4, 0.3, 0.2\}$, no player would like to participate because all possible combination are unprofitable.

In case (ii), when the aggregate marginal benefit is high enough, there might exist an equilibrium or equilibria coalitions. For example, given the set of marginal benefit is $\{0.7, 0.6, 0.35\}$, there exist two stable coalitions $\{1, 2\}$ and $\{1, 3\}$. In the former case, the internal constraint is satisfied when both players 1 and 2 have no incentive to dissolve the coalition by leaving. On the other hand, the external constraint is satisfied when player 3 has no incentive to join since the reward of free-riding is better than that of participation.

A special case of homogeneous countries Given that all countries have the identical marginal benefit of total abatement which is in the meaningful range between $(1/N)$ to 1. When the marginal benefit γ is too large $(1 \leq \gamma)$, an IEA is unnecessary because players already have the incentive to abate fully. When γ is too small $(0 < \gamma < (1/N))$, a profitable IEA is also non-existent because all players would pollute anyway. When the marginal benefit is between $(1/N)$ and 1, there may exist stable coalitions where signatories abate and nonsignatories pollute.

The payoffs for a nonsignatory j and a signatory i with a n-member coalition are

$$\begin{aligned} \pi_j(n) &= \gamma n \\ \pi_s(n) &= -1 + \gamma n \end{aligned}$$

Any nonsignatory would take the free-riding benefit and receive a higher payoff than any signatory does.

In this membership game, each player have to decide whether or not to join a coalition. Since all participants are self-enforced, players can not reject or accept new entrants [8]. On the one hand, a nonsignatory would have a higher payoff than a signatory's. With the assumption of homogeneity, everyone would prefer to be a free-rider i.e. a nonsignatory. On the other hand, if no coalition is formed, all countries would have zero payoff. A coalition is therefore necessary to all countries.

[8] After the membership status is determined, members in the coalition act as one with the joint decision made by either the majority voting (Burger and Kolstad, 2010) or a random leadership. Theoretically, results in both cases are the same since agents are assumed to have the same preference.

As mentioned previously, the stable coalition exists when the internal and external constraints both are satisfied. The *internal constraint* (1.6) requires any signatory has no incentive to leave the coalition. The left hand side of this constraint is the payoff of being a signatory $(-1+\gamma n^*)$. It is better to be a nonsignatory (0) as a collapsed coalition in the right hand side. The absence of any signatory would lead the coalition to be unprofitable $(-1+\gamma\left(n^*-1\right)<0)$[9]. Thus we can comfortably say that n^* is the smallest integer better than the inverse of the ratio of abatement benefit to cost $(n^* \geq 1/\gamma)$.

The *external constraint* (1.7) describes that any nonsignatory has no incentives to participate in a coalition. The right hand side of the constraint is the payoff of being a nonsignatory (γn^*), which is better than that of being an extra participant $(-1+\gamma\left(n^*+1\right))$ on the left hand side. The constraint is held since all nonsignatories have no incentive to join. As discussed previously, the ratio of marginal benefit to cost, γ, is between 0 and 1. With the assumption of homogeneous players, this constraint is always satisfied.

We summarise the results so far in Table 1.1. A coalition of size n^* is stable if and only if both internal and external constraints are satisfied. If the size of the coalition is smaller than n^*, the coalition collapses and no country earns anything. When the size is (n^*+1), signatories might have the incentive to leave. Hence, n^* is the stable size for the coalition. [10] ■

[9] The player most likely to leave an IEA is the one with the *highest* payoff from abatement, because that country is being asked to make transfers to other countries which can be avoided by leaving the coalition. So the relevant marginal condition for coalition members applies potentially to all members of the IEA. The internal constraint means no signatory has an incentive to leave as long as: $-1+\sum_{i\in S}\gamma_i \geq \left(n\left(S\right)-1\right)\gamma_i \quad \forall i \in S$ where S is the set of signatories and $n\left(S\right)$ is the number of signatories. As this shows it is the signatory with the highest benefit which is most likely to wish to leave (as long as this does not destabilise the IEA). However, if the coalition without country i is unprofitable, every player's payoff becomes 0.

[10] To define the stable size, following Burger and Kolstad (2010), we need the "rounding-

Number of signatories	Signatory's payoff	Nonsignatory's payoff
0	–	0
$(n^* - 1)$	0	0
n^*	$-1 + \gamma n^*$	γn^*
$(n^* + 1)$	$-1 + \gamma (n^* + 1)$	$\gamma (n^* + 1)$
N	$-1 + \gamma N$	–

Table 1.1: Corresponding individual payoffs in a coalition

This study attempts to test the theory based on heterogeneous agents by conducting an experiment. Existing experimental studies (such as Kosfeld *et al.*, 2009) assume that all agents are identical. However, this assumption is far from the reality. Even the assumption of heterogeneity is considered by Burger and Kolstad (2010), there exist more than one equilibrium coalition in their experimental design. Though the formation of IEAs could be expected, it is not enough to predict individual decisions in the membership game by these past studies. In order to address this gap in the literature, this study considers the condition of uniqueness of equilibrium. The condition provides the existence of a unique stable n^*-member coalition where n^* is the minimum number to form a profitable coalition. By this condition, individual decisions could be predicted.

Condition 1 *(Uniqueness of equilibrium)*

Suppose all players are self-interested, when the internal and the external constraints are satisfied, there exists a unique stable n^*-member coalition if and

up" function which rounds a real number up to an integer by defining $I(t)$ as the smallest integer greater than or equal to t. With this definition, we therefore claim that the equilibrium of a coalition size is $n^* = I(1/\gamma)$. The stable size of a coalition is equal to the smallest integer greater than the inverse of the marginal benefit of abatement. Any combination which achieves this condition is a possible solution. This result implies that a higher ratio (γ) causes a smaller coalition.

only if $1 + \gamma_{n^*} > \sum_{i=1}^{N} \gamma_i$

The proof is presented in Appendix 1.1.

The condition implies that the stable coalition is unique if the absence of any single signatory cannot be replaced by the entry of all nonsignatories. The unique equilibrium condition ensure that the formation is the only one profitable coalition ($-1 + \sum_{i=1}^{n^*} \gamma_i > 0$). If any player from player 1 to n^* leaves the coalition, there is no substitution to form a profitable coalition. Connecting the internal constraint ($\sum_{i=1}^{n^*} \gamma_i > 1$) with the unique equilibrium condition, we have

$$\sum_{i=1}^{n^*} \gamma_i > 1 > \sum_{i=1}^{n^*-1} \gamma_i + \sum_{j=n^*+1}^{N} \gamma_j$$

By subtracting $\sum_{i=1}^{n^*-1} \gamma_i$ from both sides, we derive that

$$\gamma_{n^*} > \sum_{j=n^*+1}^{N} \gamma_j$$

Whilst we acknowledge this indeed a strong condition, however, in order to identify the individual incentives to participate in the coalition, such a condition provides an environment where each agent has a weakly dominant strategy in terms of their own payoffs.

The following 3-player example helps us to understand the purpose of this condition.

Example of a 3-player game

Given a 3-player game, let players 1, 2 and 3 have various abatement parameters γ_1, γ_2 and γ_3 respectively[11]. There are $2^3 - 4 = 4$ possible coalition sets

[11] We define $0 < \gamma_3 < \gamma_2 < \gamma_1 < 1$

Player 1	Player 3	Payoff if Player 2 join	Payoff if Player 2 not join
IN	IN	$\max[\gamma_1 + \gamma_2 + \gamma_3, 1]$	$1 + 2\gamma_2$ when $\gamma_1 + \gamma_3 \geq 1$ 1 when $\gamma_1 + \gamma_3 < 1$
IN	OUT	$\max[\gamma_1 + \gamma_2, 1]$	1
OUT	IN	$\max[\gamma_2 + \gamma_3, 1]$	1

Table 1.2: Payoff table for Player 2

which include the full coalition set $\{(1, 2, 3)\}$, and the two-member coalition sets $\{(1, 2)\,; (1, 3)\,; (2, 3)\}$.

Table 1.2 lists the possible payoffs for player 2. The first and second column show the membership status of player 1 and 3 respectively. If player 2 decides to join the coalition, payoffs for the three possible cooperation combinations are shown in the third column. If player 2 decides not to join, the possible payoffs are listed in the fourth column.

Following the internal constraint, player 2 would form a coalition with player 1 if $\gamma_1 + \gamma_2 \geq 1$. Meanwhile, with the external constraint, player 3 has no incentive to participate if $\gamma_1 + \gamma_2 + \gamma_3 \leq 1 + 2\gamma_3$. Thus, the 2-member coalition $\{(1, 2)\}$ is a stable equilibrium. However, there could be another equilibrium $\{(1, 3)\}$ when it is also profitable ($\gamma_1 + \gamma_3 \geq 1$). If the equilibrium set has more than *one* combination, the individual incentive to participate in the coalition is not clear. Both players 2 and 3 have the incentive to cooperate with player 1, but also want to free-ride.

With the unique equilibrium condition ($\gamma_1 + \gamma_3 < 1$), player 3 has no incentive to cooperate with others. Hence, joining is the dominant strategy for both player 1 and 2.

Figure 1.1 presents the marginal benefits to three different players into three dimensions. The parameters are ranked from high to low as $\gamma_1 > \gamma_2 > \gamma_3$. The

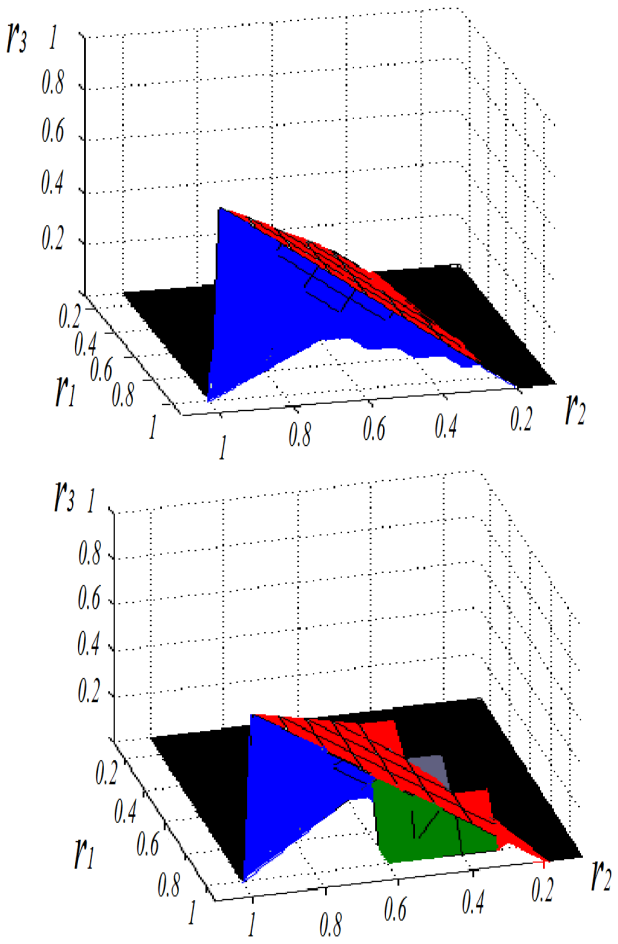


Figure 1.1: Equilibrium boundary in a 3-player game

internal and the external constraints are the blue area and the red one in the figure respectively. When the parameter sets are within the constraints (which is in the middle of the chart), there exist stable coalitions, anywhere beyond the boundary is an unprofitable coalition set. There might exist multiple equilibria in this space. Taking the set $\{0.8, 0.7, 0.6\}$ for an example, the coalition sets $\{1, 2\}$, $\{1, 3\}$ and $\{2, 3\}$ are also stably profitable. If the parameters are too small to be in the space, the coalitions are unprofitable and unstable (which is in the right black area of the chart). Taking the set $\{0.3, 0.2, 0.1\}$ for example, any combination from this set cannot form a profitable coalition.

As mentioned earlier, this study focuses on a unique equilibrium only. The boundary of the unique equilibrium condition is presented as the green area in the lower chart. The space within the areas of blue, red and green is where a unique equilibrium could exist. The marginal benefit of player 3 is not large enough to encourage the player forming a coalition with either player 1 or 2. In other words, both players 1 and 2 are irreplaceable by player 3. Compared to the space in the upper figure, the space in the lower figure is smaller. Though our focus is subject to specific set-ups, individual decisions are still easier to be predicted and explained through this experimental analysis. ■

1.2.2 Inequality-averse preference in a coalition game

The constraints above are considered assuming individuals have egoistic preferences. As mentioned previously, this assumption fails to capture the idea that individuals may behave differently in a practical interactive game. In order to address this limitation, we now incorporate the idea of "other-regarding" preferences into our analysis to examine individual incentives.

Following Fehr and Schmidt (1999), we assume that subjects dislike unfair outcomes at different levels. Subjects feel disadvantaged when they are better off or worse off in material terms. With this concept, the utility of a player k of a profitable n-member coalition can be represented as

$$\begin{aligned} & u_k(n) \qquad (1.8) \\ = \; & \pi_k(n) - \frac{\alpha_k}{N-1} \sum_{k' \neq k} \max\left(\pi_{k'}(n) - \pi_k(n), 0\right) - \frac{\beta_k}{N-1} \sum_{k' \neq k} \max\left(\pi_k(n) - \pi_{k'}(n), 0\right) \end{aligned}$$

where Player k' denotes all players except player k. The first term is the payoff of player k and the second term indicates the average utility loss from other player k' with the disadvantage-loss parameter α_k. The third term measures the average loss from other player k' with the advantage-loss parameter β_k, which is assumed within the range between 0 (inequality-neutral) and 1 (highest degree of inequality-aversion).

Extended from the constraints and Condition 1, an unique n^*-member coalition exists when all agents are self-interested. When the individual inequality-aversion is considered in the utility function, the following hypothesis provides the conjectured outcome of coalition formation.

Conjecture 2

If the internal and external constraints hold and the condition for the unique equilibrium is satisfied, the coalition formation could be either a unique n^*-member coalition, or a unique coalition which is larger than n^*, or an unstable coalition with different inequality-averse preferences.

The explanations of the possible outcomes are shown in Appendix 1.3.

Three possible outcomes are depending upon different circumstances of individual inequality-averse preferences :

(i) When all players have no inequality-aversion or a low degree of inequality-aversion, there exists a unique n^*-member coalition equilibrium.

(ii) When any player from players $n^* + 1$ to N has a high degree of inequality-aversion (large β), the external constraint could be violated. If other things are equal, the coalition formation is stable and larger than n^*.

(iii) When any player from players 1 to n^* has a high degree of inequality-aversion, the internal constraint could be violated. The coalition formation then becomes unstable.

Without taking inequality-aversion into account, a unique stable coalition is formed with three constraints. When the inequality-aversion is considered as part of the individual preferences, there are a number of effects. First, inequality-aversion reduces countries' utility when payoffs are not equal. The incentive of being a nonsignatory therefore decreases and the external constraint is more likely to be violated. This will tend to increase the size of a stable coalition.

Second, countries with strong inequality aversion would be encouraged to stay in an IEA or join it to spread the benefits of equalisation because of the transfer mechanism where signatories share the same coalition payoff. However, except for a grand coalition, any combinations of IEAs has a free-riding effect. An expanding IEA will tend to exacerbate the payoff gap between signatories and nonsignatories. Signatories with a strong sense of inequality-aversion may violate the internal constraint if the payoff gap is large. Under this condition, the most likely outcome would be to have no IEA at all, so a certain level of inequality aversion can destabilise an IEA.

When inequality-aversion is taken into account, the net effect of these two factors shapes the stability and the formation of IEA. When a country decides to join a coalition given the first effect, the participation will lead to a smaller advantage loss but a larger disadvantage loss. With this character, a stable coalition can not be easily expand by the first effect. On the other hand, as long as stable equilibrium is not a grand coalition, there exists inequality. The payoff gaps between signatories and nonsignatories are enlarged with the second effect. The internal constraint is more difficult to be satisfied and the coalition formation becomes unstable.

The following example could improve our understanding.

A Numerical example

Here is a numerical example to explain this proposition. Supposed that there are five agents with various marginal benefits of total abatement, (0.675, 0.375, 0.125, 0.1, 0.075). When the agents have no inequality-aversion or a low degree of inequality-aversion which is no more than 0.4, agents 1 and 2 follow the internal constraint to join the coalition while agents 3, 4, and 5 follow the external constraint and stay away from the coalition. The formation of the coalition would therefore converge to the 2-member coalition equilibrium. The coalition is stable and profitable over a 100-round repeated game. The total contribution in 100 rounds is presented in the upper chart in Figure 1.2.

When the internal constraint is violated due to an agent having a degree of inequality-aversion higher than 0.4, the coalition is no longer stable. The lower chart in Figure 1.2 shows the case where agent 2 violates the internal constraint when his inequality-aversion factor α is greater than 0.4. This is a 'noisy' result is due to the high degree of inequality-aversion of agent 2.

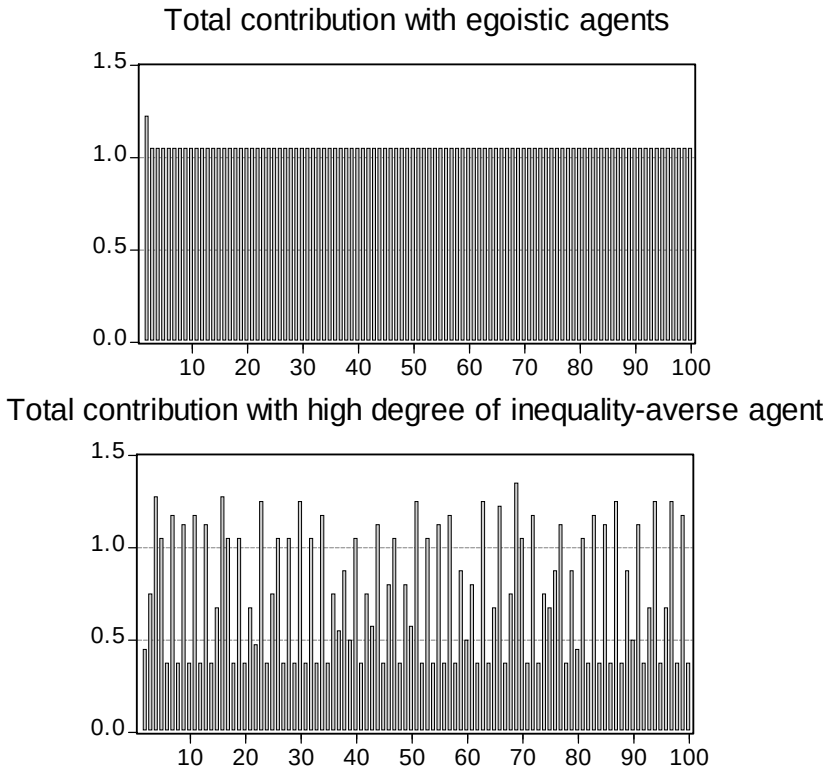


Figure 1.2: Numerical example of a 5-player coalition game

The agent only has an incentive to join when the coalition size is large enough. Nevertheless, other nonsignatories have no intention of giving up the free-riding benefit and participating in the coalition. Hence, the consequence is that the coalition is unstable over rounds. ■

In terms of the design of this particular example, the external constraint will not be violated given the highest degree of inequality-aversion. Hence, a larger stable coalition is not possible in this case.

1.3 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). There were 50 subjects who were registered on the ORSEE registration system by Greiner (2004). They were students from different countries and in various disciplines at the University. This sample that mimics the diversity in the real world where international policy makers and multidisciplinary knowledge are present helps understand IEAs formation. The instructions (see the Appendix 1.4) were provided on subjects' desks. The instructions consist of three parts. This chapter endeavours to investigate the coalition formation through individual preferences of inequality-aversion. The data are drawn from part 1 and part 3 in the experiment.

To ensure the data quality, the subjects had to comprehend the rules of the game as much as possible. To do so, the experimenter introduced the rules and gave the participants time to read through the instructions thoroughly and accomplish the controlled questions. In 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.

According to our assumption, the subjects should be self-motivated. The subjects were therefore required to maximise their own payoffs. In addition, to simplify the experiment, the subjects were not allowed to exchange information; no conversation was allowed (except for asking the experimenter to clarify the questions) during the experiment.

A questionnaire was circulated before the experiment to gather demographical information, including the subject's degree disciplines, age (the year they were born), ethnicity, political orientation, and the level of belief in a religion. This questionnaire is designed to gather more explanation on their decision-making in the experiment. The first three questions are objective and the data shows the diversity of the participants. The results are presented in the Figures 1.3, 1.4, 1.5, and 1.6. Figure 1.3 shows subjects' major: 11 participants recruited were reading Economics; 8 participants in Humanities; 13 participants in Science; 1 participant in Laws; 9 participants in Engineering; 1 participant in Psychology; 7 participants in other disciplines and no recruit was reading Business-related disciplines. Figure 1.4 shows the distribution of ethnicity: 32 subjects were white; 15 were Asian or Asian British; 2 were Black or African or Caribbean or Black British; and 1 fell into the category of any other ethnic groups. Also, all participants were undertaking undergraduate or postgraduate courses at the University and their average age was 25 years-old (the oldest being 45 and the youngest being 21).

The last two questions were concerning their subjective preferences. Figure 1.5 presents the distribution of their level of belief on religions while subjects

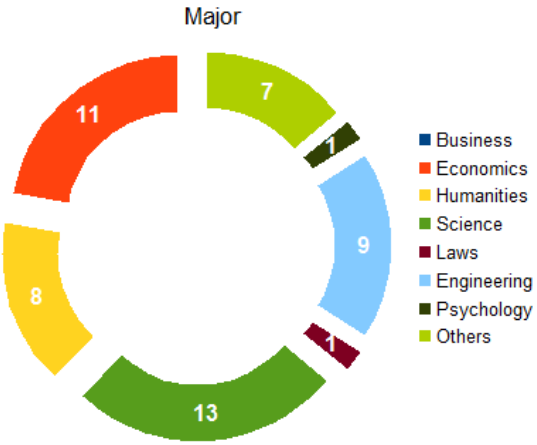


Figure 1.3: Degree subject distribution

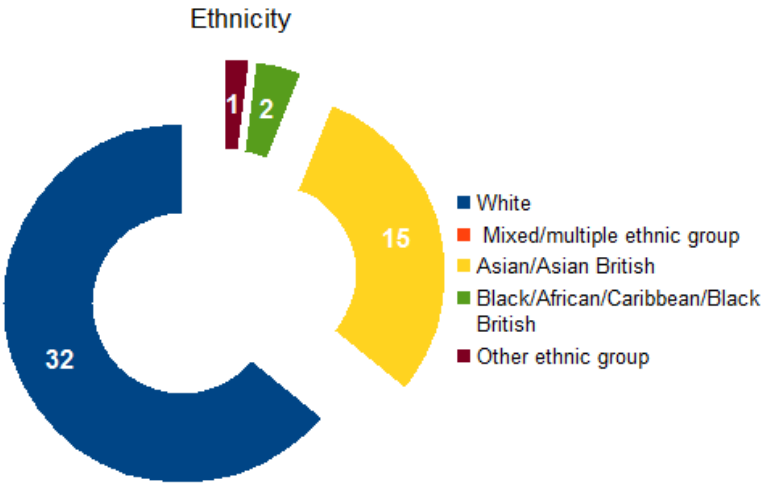


Figure 1.4: Ethnicity distribution

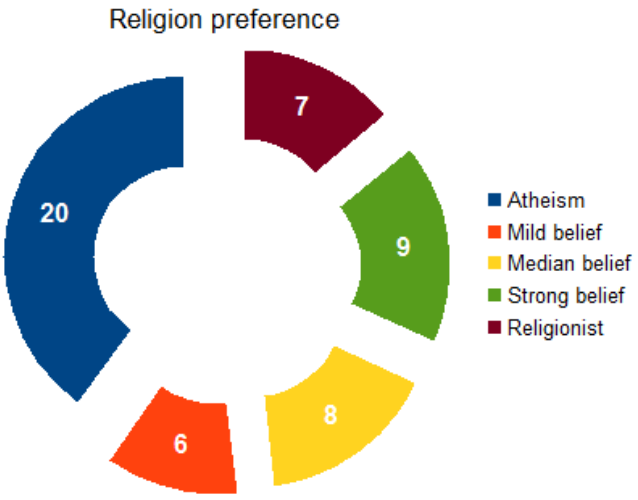


Figure 1.5: Religious preference distribution

were asked to identify themselves on a scale ranging from level 1 (not religious at all) to 5 (extremely religious). In the results, 20 subjects consider themselves to be atheist. Meanwhile, 6, 8, 9, and 7 subjects consider themselves as religious, with mild belief, median belief, strong belief and pure religionists respectively. The average level is 2.5. The distribution shows that the subjects' religious belief is between mild to median belief, overall.

The other question aims to indicate the subjects' political preference (level one indicates left, level two centre-left, level three neutral, level four centre-right and level 5 right). The distribution is presented in Figure 1.6. In our sample, 7 subjects self-identified themselves as left wing; 10 as centre-left; 25 as neutral; 7 as centre-right and 1 as right wing.

The main experiment is comprised of two parts, as shown in Parts 1 and 3 in Appendix 1.4. The experimental procedure was designed as follows.

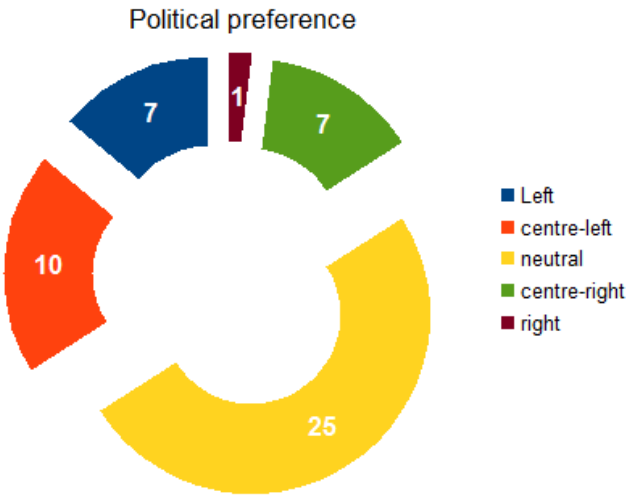


Figure 1.6: Political preference distribution

1.3.1 An inequality-averse preference test

The instructions of the first part are shown at Part 1 in Appendix 1.4. This test aimed to examine the subjects' individual attitude towards inequality-aversion. To measure a subject's inequality-averse preference, the subjects who did not know each other were paired together. **The subjects did not know their partners and the partners' decisions during the whole test.** Their payoffs were determined by their own decisions as well as their partner's decisions. This was to understand the individual preferences without knowing their strategies they played. The subjects were required to answer a series of decision questions in 11 rounds as shown in Table 1.3. Option 1 meant the subjects share the same allowance, while Option 2 meant the subjects could take all-or-nothing with a certain probability.

Given the allowance £5, which would be shared by a subject (denoted as A afterward) receiving x and another subject (denoted as B afterward) receiving

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

Table 1.3: Distribution of payoff in all 11 rounds in the inequality-aversion test

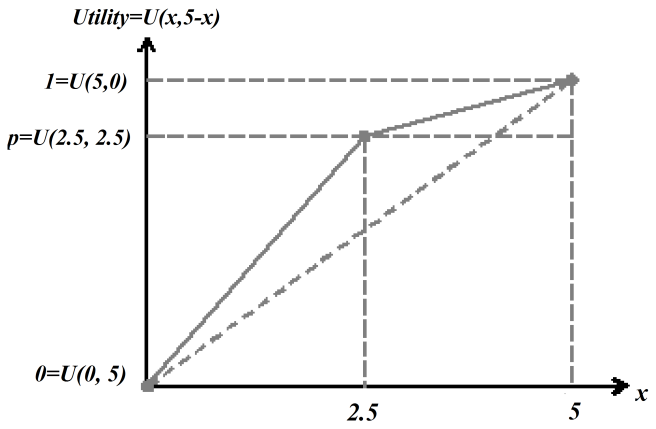


Figure 1.7: Subject A's inequality-averse preference

$(5-x)$. Subject A's inequality-averse utility was determined by both her and the other subject's shares as displayed in Table 1.3.

$$U_A(x, 5-x) = \begin{cases} x - \alpha[(5-x) - x] & \text{if } x \leq 2.5 \\ x - \beta[x - (5-x)] & \text{if } x \geq 2.5 \end{cases} \tag{1.9}$$

The upper function represents Subject A's utility when A has less than half of the total allowance. The parameter α is the coefficient of the average disadvantage loss of A. On the other hand, when A has more than half of the total allowance, the lower function is A's utility with the coefficient of the average advantage loss.

The function can be presented as the solid line in Figure 1.7. The horizontal axis is the allowance of A while the vertical axis is A's corresponding utility.

The utility depends on the payoff set of subject A and the opponent B which is presented as (£x, £$5-x$). From (1.9), we derive that A 's utility of

(£5, £0) is $U_A(5,0) = 5 - 5\beta$, and the utility of (£0, £5) is $U_A(0,5) = -5\alpha$, and the utility of (£2.5, £2.5) without any inequality is $U_A(2.5, 2.5) = 2.5$. We normalise by setting $[U_A(5,0) - U_A(0,5)]/5 \equiv 1$.

Given that a series of probabilities is involved in the inequality test, this test could be characterised by strategic uncertainty. The subjects' risk attitudes may be involved in their decisions. For instance, even the expected payoff of taking Option 2 is higher than the payoff of Option 1, a risk averse subject may prefer to the equal-share option because she or he fears the possible loss by taking Option 2. There are some experimental designs, such as Blanco *et al.* (2011) and Yang *et al.* (2012), that attempted to exclude strategic uncertainty. and avoid risk attitudes. They employed two games to capture the factors that advantage or the disadvantage the subjects.

The relationship between risk-aversion and inequality-aversion has been discussed by several recent studies. An experimental study by Carlsson *et al.* (2005) also found that people who are inequality-averse are more risk-averse, and that the reverse relation also holds true: risk-averse individuals tend to be more inequality-averse. Given the same individual risk, Kroll and Davidovitz (2003) provided another experimental evidence that most of the subjects preferred equal distribution to inequality.

Whilst it should be noted that our experimental design did not exclude the subjects' risk attitudes, our design is still superior in the sense that the normalisation provides a normalised inequality-averse utility in one game[12]. While other studies avoid strategic uncertainty in their experiments, there exist

[12]We acknowledge that there are other methods to measure attitudes to inequality. Different to other experiments focus on social preferences, there were two social preferences tests and one public good game in our experiment. This design could measure individual inequality-averse attitude without complex procedures.

other factors which could lead to a biased estimation of inequality-aversion. For example, Yang *et al.* (2012)'s experiment shows that subjects may have a negative advantage loss. It implies that subjects may prefer to show off rather than feel guilty when they are advantaged. Such bias does not arise in our design because the utility has been normalised.

To find out the inequality-averse preference, we asked each subject to choose between two options in each row of Table 1.3. The first option is a certain option where both players share the allowance equally (£2.5). The second option is an uncertain option that the subject would win all-or-nothing depending on probability. The given probability decreased by 10% in each round.

Since the allowance was a good, the subjects in theory would prefer to have more. The first row in Option 2 shows that if the probability to yield (£5) is 1, any subject would choose Option 2 rather than Option 1. On the other hand, at the bottom row in Option 2, if the probability of the set (£5, £0) is equal to 0, subjects would prefer Option 1 rather than Option 2. Hence, we assume that subjects will choose Option 2 in the first few rows and Option 1 in the last few. For each subject with a consistent preference, there exists a point with a certain probability where the subject would switch from Option 2 to Option 1. We denote the probability of (£5, £0) at the switch point by p. Then subjects feel indifferent between (£2.5, £2.5) for sure and (£0,£5) with probability $(1-p)$ and (£5, £0) with probability (p). Such probability p can be seen as the weight of inequality aversion.

In Option 2, a subject is given (£5) with the probability p and (£0) with the probability $(1-p)$. In Option 1, the subject is given (£2.5) for sure. The subject would feel indifferent between the sharing combination (£2.5, £2.5) and the mixed combination of (£0, £5) with probability $(1-p)$ and (£5, £0)

with probability (p). We can present this in an equation as

$$U\left(2.5, 2.5\right) = \left(1 - p\right) U\left(0, 5\right) + pU\left(5, 0\right) \tag{1.10}$$

The inequality-averse parameters α and β would be found through p. Given that the range between the utility of all $U\left(5, 0\right)$ and nothing $U\left(0, 5\right)$ is normalised, the inequality-averse preference was indifferent when subjects are disadvantaged and advantaged ($\beta = \alpha$). Although it was mentioned earlier that a player might suffer more from inequality when she is disadvantaged ($\beta \leq \alpha$), there are two reasons that support us to do so. In practice, it is not easy to find a subject's preference without standardising the unit of the utility. In the literature, the experimental evidences show that the disadvantage factor is not necessarily smaller than the advantage factor (Dannenberg *et al.*, 2007; and Yang *et al.*, 2012).

Hence, we assume that the inequality-averse preference are indifferent to being disadvantaged and advantaged.

When the subject is advantaged, $U(5, 0)/U(2.5, 2.5) = 1/p$, we have

$$\alpha = \beta = p - \frac{1}{2} \tag{1.11}$$

Since the probability p is in the range of 0 and 1, the inequality-averse parameters α and β are at the range of $-\frac{1}{2}$ to $\frac{1}{2}$.

Subjects are inequality-neutral when their switch points are at $p = 0.5$ where the expected payoff is equal to the fair payoff. The inequality-averse preference $\alpha = \beta = 0$. In other words, the utility of taking all the allowance (£5) is not two times higher than that of equally sharing the allowance (£2.5).

When the switch point is $p > 0.5$, subjects are inequality averse and their utilities are lower than their monetary payoffs. The extreme case is when $p = 1$, and β is 0.5. It implies that subjects have indifferent preferences of taking one unit payoff or equally sharing the allowance. When the advantage aversion is very high ($\beta > 0.5$), it is considered as altruism, which is not able to capture in this design[13]. Altruists would prefer to give goods to others in order to achieve fairness. Although it is beyond the scope of this study, altruism is an important topic that needs to be explored in future studies as it can happen in reality.

When p is less than 0.5, subjects are not inequality-averse (neither advantage acceptors nor disadvantage acceptors). While Fehr and Schmidt (1999) exclude inequality acceptors in their assumption, inequality-aversion is considered in this study as it may happen in the experiment. For these subjects, they would be considered as inequality-lovers or risk-lovers (because the experiment has strategic uncertainty). Both inequality and risk lovers are possible but uncommon in reality (as seen in the experimental result later), so our study does not focus on this issue. Hence, these subjects have been excluded from our sample[14].

1.3.2 Experiment of a coalition game

The instructions of this part are shown in Part 3 in Appendix 1.3.

To concentrate on the entry decision, we simplify the two-stage game into the membership game. The scenario in the second stage has been modified

[13]Because the probability p is only in the range of 0 and 1.

[14]The existing probabilities in the test may introduce a bias by involving risk-averse preference and hence weaken the conjecture.

to show the situation when a *profitable* n-member coalition is formed (the coalition generates a positive payoff if the aggregate benefit-to-cost ratio of signatories is larger than 1, $\sum_i^n \gamma_i > 1$). In this case, all signatories abate and all nonsignatories pollute. Otherwise, the coalition *collapses* and all players pollute. Hence, all elements in the payoff set $(\pi_1(n), \pi_2(n), ..., \pi_N(n))$ are non-negative. It implies that all players behave rationally in maximising their payoffs.

The social welfare is the aggregated payoffs from all nonsignatories and the coalition payoff . The maximum welfare exists when the grand coalition is formed [15]. All players face a dilemma of being a nonsignatory with free-rider payoff or being a member with the shared payoff.

A public good game with various payoff tables was conducted. The results from the previous part were used to predict whether the subjects would violate the stability constraints in the coalition game. In the theoretical model, this is a two-stage game. The first stage is the membership game, where subjects decide whether or not to join a coalition. The second stage is the abatement game. In the abatement game, since the payoff is a linear function, the decision-making would be straightforward. When a subject decides to join a coalition, she would abate fully at the second stage. When her decision is not to join, she would not abate at all at the second stage. Based on this, we simplify the two-stage model to a one-stage membership game in the experiment.

In this part, subjects were randomly assigned to groups of five persons. They did not know who they were playing with, but they did know that they

[15]The total payoff is $\Pi = \Pi_s + \sum_j \pi_j = [(-n) + \sum_i^n \gamma_i] + \left[n \sum_j \gamma_j\right]$. Because only a profitable coalition is counted, the total payoff is maximised when the grand coalition is formed $\Pi = (-N) + \sum_i^N \gamma_i$.

were playing with the same people during the whole session. In our assumption, subjects should be self-motivated. Subjects were therefore required to maximise their own payoffs.

In each treatment, each subject was given a particular payoff table of all the possible coalition combinations. A group of N subjects would generate $\left(2^N - N - 1\right)$ combinations. In order to generate a simple and clear table for subjects, the number of 5 subjects was set in a group with 26 possible combinations.

The game was a one-shot game, and decisions in each round were independent.

With this design, the subjects know no more than their own inequality-averse preference. However, the experiment allowed subjects to have a learning process so that the coalition would converge to the Nash equilibrium. The game was played 15 times in each sub-treatment. Subjects were given 180 seconds to make their decisions of whether or not to join the coalition. According to the pilot experiment, this time setting gave subjects enough time to make their decisions. Any decision which was not made within this amount of time would be counted as non-participation. This rule is sensible because the decision was asked whether or not to join a coalition with a non-participating status.

Finally, the coalition formation and all subjects payoffs in the group were reported on the result screen.

Subjects should make their decisions based on their economic incentive. In order to ensure subjects were aware of their profit-maximising incentives rather than other non-economic incentives, the reference to environmental issues was removed from the instruction. The level of marginal benefit of the

Round	Player 1	Player 2	Player 3	Player 4	Player 5
$1-15$	0.675^*	0.375^*	0.125	0.10	0.075
$16-30$	0.075	0.15^*	0.25^*	0.3^*	0.35^*
$31-45$	0.40^*	0.65^*	0.075	0.10	0.125
$46-60$	0.05	0.1	0.4^*	0.35^*	0.3^*
* means the weakly dominant strategy of the player is joining the coalition.					

Table 1.4: List of parameters of marginal benefit for players taking Treatment 1

Round	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 the weakly dominant strategy of the player is joining the coalition.					

Table 1.5: List of parameters of marginal benefit for players taking Treatment 2

total abatement was labelled as parameter (γ_k, $\forall k \in [1, ..., 5]$) in the experimental design. There are two treatments with different parameter sets. 20 subjects took Treatment 1 and the rest of the subjects took Treatment 2. The individual parameters in the Treatment 1 are listed in Table 1.4, and the parameters in Treatment 2 are listed in Table 1.5.

According to Condition 1, we can claim that a unique equilibrium could be found in some particular cases. The theoretical result suggests that a unique equilibrium exists within the internal, the external and the unique constraints. To achieve a unique equilibrium, the experiment was built with some particular parameters mentioned earlier in the theory. Subjects with high marginal benefit parameter are labelled (*) in Tables 1.4 and 1.5, they were predicted to have

a weakly dominant strategy to *join*. Eight treatments within the constraints were selected in the experiment. The theoretical size of the stable coalition in treatments was from 2 to 4. Each group was given four sub-treatments with a different number of subjects predicted to be in the stable coalition.

Tables 1.4 and 1.5 present the treatments which were designed to ensure a unique stable IEA based on the assumption of no inequality-aversion. Each sub-treatment had a unique equilibrium and each subject had a weakly dominant strategy in the membership game. Meanwhile, we propose in Conjecture 2 that different attitude to inequality-aversion may lead to higher membership or no stable IEA. The internal constraint is more likely to be violated by individuals with high degree of inequality-aversion. But due to the internal transfers, a nonsignatory would gain less advantage loss but more disadvantage loss if she or he decides to join a coalition. Hence, the external constraint is not easy to be violated. The experiment in this study is able to test whether subjects with high inequality-aversion are more likely to violate the internal constraint and lead to unstable.

Given the particular parameter, each subject was assigned an individual payoff table which contained all possible coalition combinations with the corresponding payoffs. If the possible coalition was profitable, from (1.5), the payoff of a subject who decided to join is

$$\pi_s = \begin{cases} 30 \times (-1 + \sum_i^n \gamma_i) & \text{when } \sum_i^n \gamma_i \geq 1 \\ 0 & \text{when } \sum_i^n \gamma_i < 1 \end{cases}$$

Meanwhile, from (1.2), the payoff of a subject who decide not to join is

$$\pi_j = \begin{cases} 30 \times (n \times \gamma_j) & \text{when } \sum_i^n \gamma_i \geq 1 \\ 0 & \text{when } \sum_i^n \gamma_i < 1 \end{cases}$$

The monetary payoffs were 30 times higher than the theoretical payoffs in the previous section. This design did not affect the theoretical predictions, but the diversity of the marginal benefits became more significant to subjects.

When a possible coalition is unprofitable, all subjects in the group gain nothing for return. The possible payoffs for subjects were from £0 and up to £24. The payoff depended on the given parameters and the coalition formation. In the experiment, we simplified the decision-making process by reducing the calculation process. With the payoff table, subjects could easily find the corresponding possible payoffs without working on the payoff function.

Given the results obtained in the inequality-averse test, an example at Table A1-1 in Appendix 1.4 is explained as follows. The table illustrates 26 possible coalition combinations for 5 players [16] and the corresponding payoffs. A stable coalition is formed when the internal and the external constraints are held. An unique stable 3-member coalition exists when Players 3, 4 and 5 obey the internal constraints and Players 1 and 2 obey the external constraints.

In the case of the external constraint, we assume that all subjects are inequality neutral except for Player 1. Player 1 would obey the constraint if the utility of being a nonsignatory $(6.75 - \frac{2.25}{4}\alpha - \frac{15.75}{4}\beta)$ is better than being a signatory $(3.75 - \frac{8.25}{4}\alpha)$. However, the subject would violate the external constraint when she has high inequality aversion. Since the disadvantage-aversion is indifferent to the advantage-aversion, Player 1 would violate the

[16] A possible coalition combination requires at least 2 players. Thus the number of the possible coalition combinations is $2^5 - (5 + 1)$.

external constraint when $\frac{16}{13} < \alpha$ (or $p > \frac{45}{26}$) . However, altruism cannot be captured in this test because Player 1 is unlikely to join the coalition with Players 3, 4 and 5, as mentioned earlier.

Similarly, Player 2 would violate the external constraint only when the subject's preference $p > \frac{37}{26}$. It means that Player 2 is very unlikely to join the coalition.

In the case of the internal constraint, if others are inequality-neutral, Player 3 would follow the internal constraint when the utility of joining $(1 - \frac{8.5}{4}\alpha)$ is higher than the utility of not joining (0). However, if Player 3 has strongly inequality-averse preference, $p > 0.97$, Player 3 would violate the internal constraint and not join the coalition. With the unique coalition condition, whether the external constraint is obeyed by others or not, the equilibrium would be a failed coalition because Players 3, 4 and 5 are irreplaceable.

Similarly, Players 4 and 5 would violate the internal constraint if their preference $p > 0.97$.

We can therefore calculate the threshold to break the internal and external constraints for each subject. Subjects who break the external constraint would have very high advantage aversion. However, we should note that altruism can not be captured in our test. On the other hand, the internal constraint is more likely to be violated. The thresholds are also very high. This could explain that subjects are likely to follow their weakly dominant strategies.

1.3.3 The results from the experiment

In the inequality-averse test, each subject was asked to choose from two options in 11 rounds. In the theoretical prediction, the decision in round 1 would be

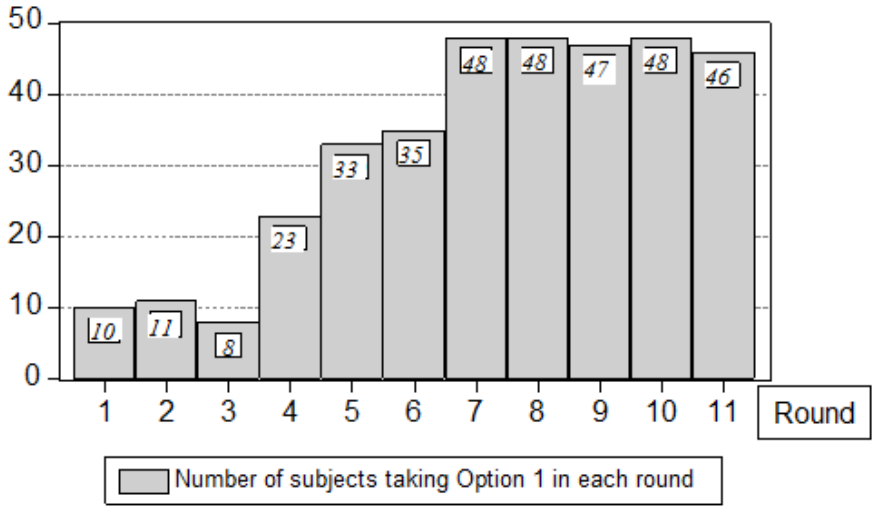


Figure 1.8: Number of subjects taking 'Option 1' in each round

'Option 2' and the decision in round 11 would be 'Option 1'. One turning point was expected and that was when the decision changed from Option 2 to Option 1. The result demonstrates that 33 out of 50 subjects had no more than one switching point in 11 rounds, while 2 subjects took Option 1 in the whole part. The degrees of inequality-aversion were therefore determined.

Figure 1.8 presents the number of subjects taking Option 1 in each round. The majority had their switch point at rounds 3, 4, 5, or 6. After round 7, almost every subject took Option 2. Although the experimental design allowed 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 earlier, five subjects were excluded because they were negative inequality-averse.

Table 1.6 shows the OLS estimation of inequality-averse preference. The dependent variable is average times of taking the Option 1 in the inequality-

Variable	Inequality-aversion level OLS Regression		
Constant term	−12.53 (11.15)		
AGE	0.007 (0.006)		
POLITIC	0.005 (0.03)		
RELIGION	−0.02 (0.02)		
Log Likelihood	19.13514	R-squared	0.042
Total Observation	50		

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

Table 1.6: OLS estimation of inequality-averse preference

averse test. Independent variables are subjects' age (AGE), political attitude (POLITIC), and religious attitude (RELIGION). The result shows that these factors from our questionnaire have insignificant effect on subjects' inequality-averse preferences.

In the membership game, all subjects were put into 10 groups and took four sub-treatments in 60 rounds. Groups 1 to 4 used Treatment 1 in Table 1.4 and groups 5 to 10 used Treatment 2 in Table 1.5. Each subject in the group was given a different value of the marginal benefit parameter γ. This parameter implied their contribution to the group, if they decided to join in. When the total contribution of a group was over 1, the coalition was profitable and everyone received the payoff which depended on their decisions. Otherwise, an unprofitable coalition brought nothing to all the players in the group. With the assumption of no inequality-aversion, the peculiar design of this ex-

periment leads to a unique equilibrium and the total contribution of this stable coalition is 1.05.

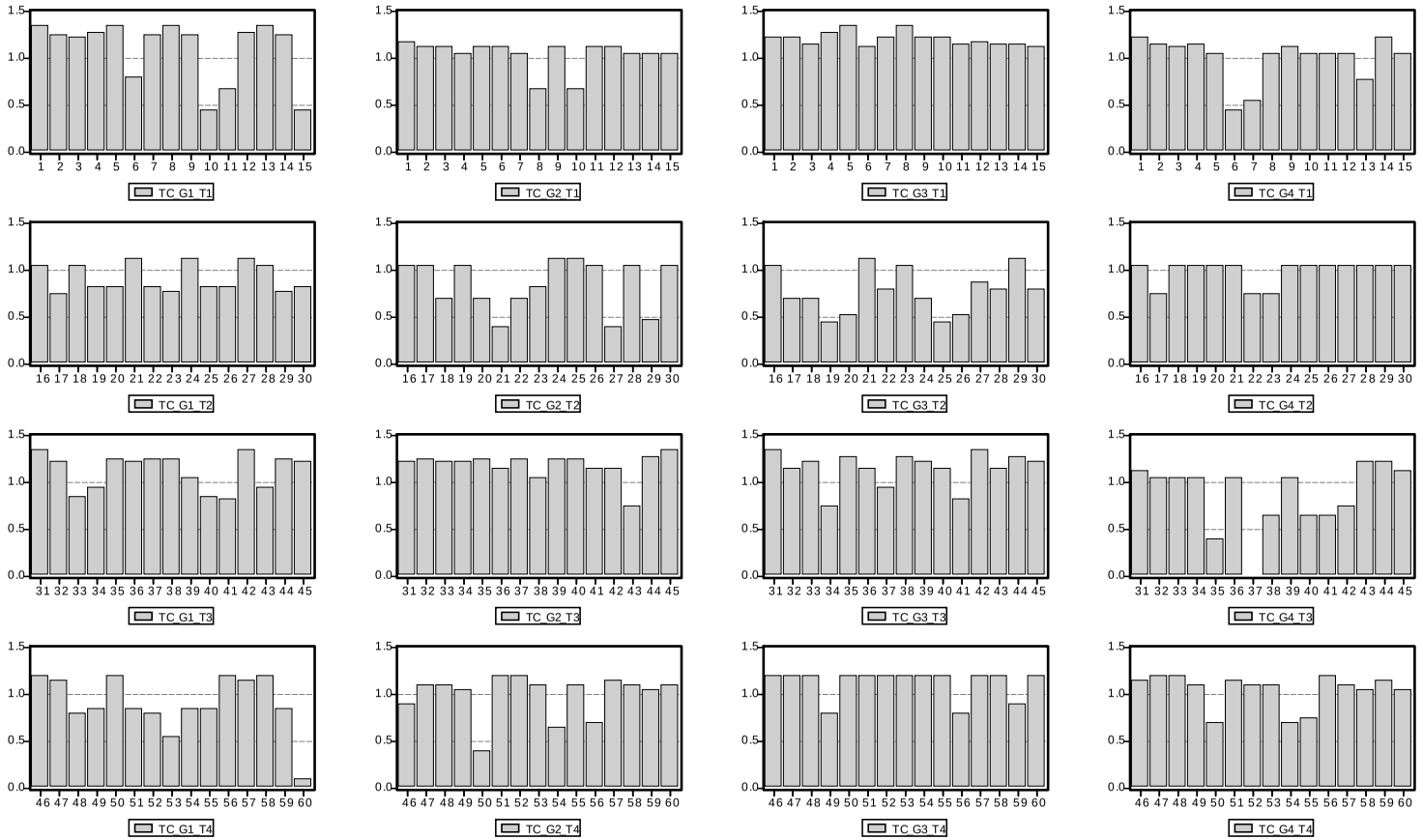


Figure 1.9: The total contribution of Group 1-4 in four sub-treatments

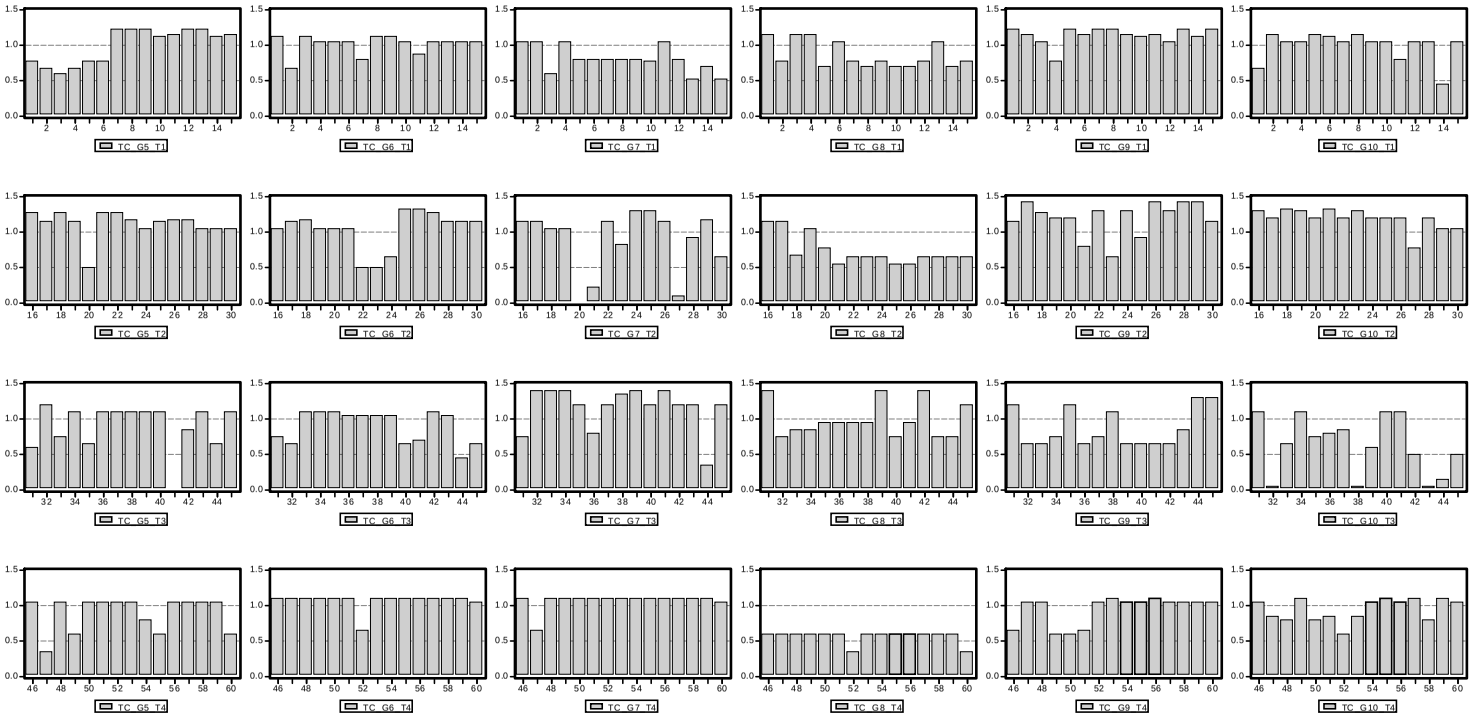


Figure 1.10: The total contribution of Group 5-10 in four sub-treatments

Figure 1.9 shows the results of the total contribution of groups 1 to 4. The charts in the first row present the total contribution of groups 1, 2, 3, and 4 in sub-treatment 1 respectively. Similarly, the charts in the second, third and fourth rows present the total contribution of groups 1, 2, 3, and 4 in sub-treatments 2, 3 and 4 respectively.

Figure 1.10 shows the results of the total contribution of groups 5 to 10. The charts in the first row present the total contribution of groups 5, 6, 7, 8, 9 and 10 in sub-treatment 1 respectively. Similarly, the charts in the second, third and fourth rows present the total contribution of groups 5, 6, 7, 8, 9 and 10 in sub-treatments 2, 3 and 4 respectively.

In light of the study population, profitable coalitions were formed in 387 of 600 rounds. The various forms of group formation lead to different group payoffs. For example, group 6 and group 8 both take Treatment 2. Group 6 forms profitable coalitions in 47 rounds, but group 8 achieved that in only 12 rounds. Both treatments provided subjects with weakly dominant strategies to take. If subjects in the group all made their weakly dominant strategies, the internal and external constraints were held and the coalition was at Nash equilibrium. It happened in 112 out of 600 rounds and such a coalition was not stable as predicted in the theory. According to the experimental results, more than two third of the profitable coalitions were formed and they were larger than the Nash equilibrium size.

As shown in Figures 1.9 and 1.10, the formation of a coalition is neither stable nor convergent to a equilibrium in 15 rounds. Compared to the numerical example in Figure 1.2, the experimental outcome shows a similar kind of fluctuations. If the hypothesis is true, this interesting result could be inter-

preted as the effect of inequality-aversion. In other words, the inequality-averse preference has an impact on the coalition formation.

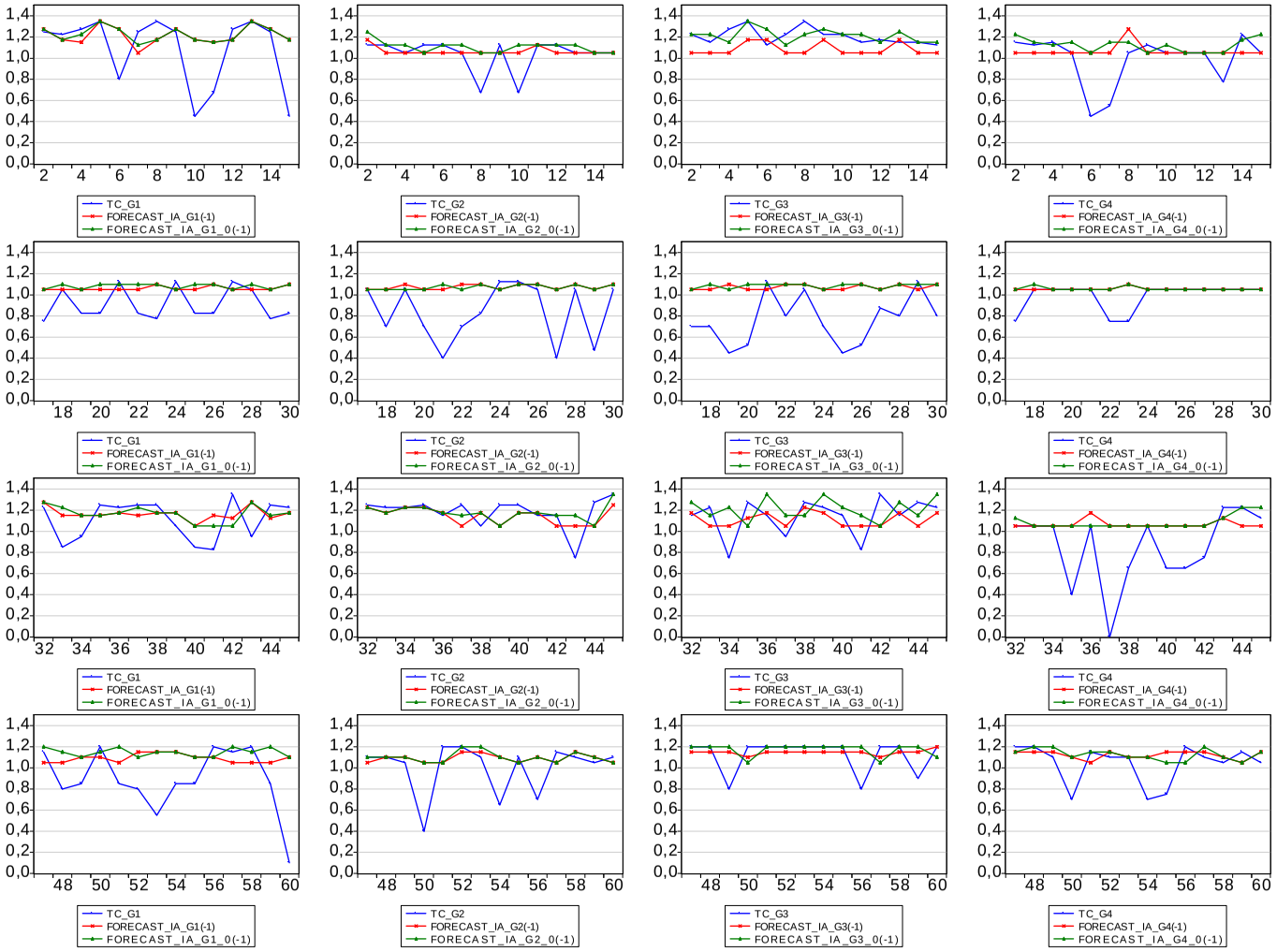


Figure 1.11: The actual total contribution and the predicted total contribution of Groups 1 to 4

In order to test our hypothesis, we examine the subjects' decision in the past round and their individual inequality-averse preferences to predict their next move. The indicated level of inequality-aversion is therefore employed to predict individual decisions in a coalition game. Figure 1.11 presents the total contribution of Groups 1 to 4. Similarly, the actual total contribution and the predicted total contribution with and without inequality-aversion of Groups 5 to 10 are shown in Figures 1.12 and 1.13.

The blue line with spots in each chart presents the actual total contribution in a sub-treatment. Given the results in the past round, the red line with cross marks are the prediction of the total contribution with the decision in the past round and subjects' individual inequality-averse preferences. There are two main reasons for employing this prediction. First, the subjects know their own inequality aversion parameter, but not others. The test in Part 1 of the experiment was anonymous and independent of Part 3, the subjects should not learn others' inequality-averse preferences. Second, learning and reciprocity are not considered in our model. Though the experiment design allows subjects finding their dominant strategy, it is not expected to figure out other's social preference. Since the subjects know no more than their own individual preferences and the historical decisions on the membership game, our prediction should be based on such information[17].

In order to examine our conjecture, the green line with triangle marks is generated only with the individual decisions in the past round only. In other words, this predictions are based on the assumption of neutral inequality-averse

[17]This experimental design attempts to purify the individual decision, any bias from other subjects' preferences should be minimised. It would be a potentially interesting but very complex issue to model (essentially testing Bayesian learning), we will leave this challenge to the future studies.

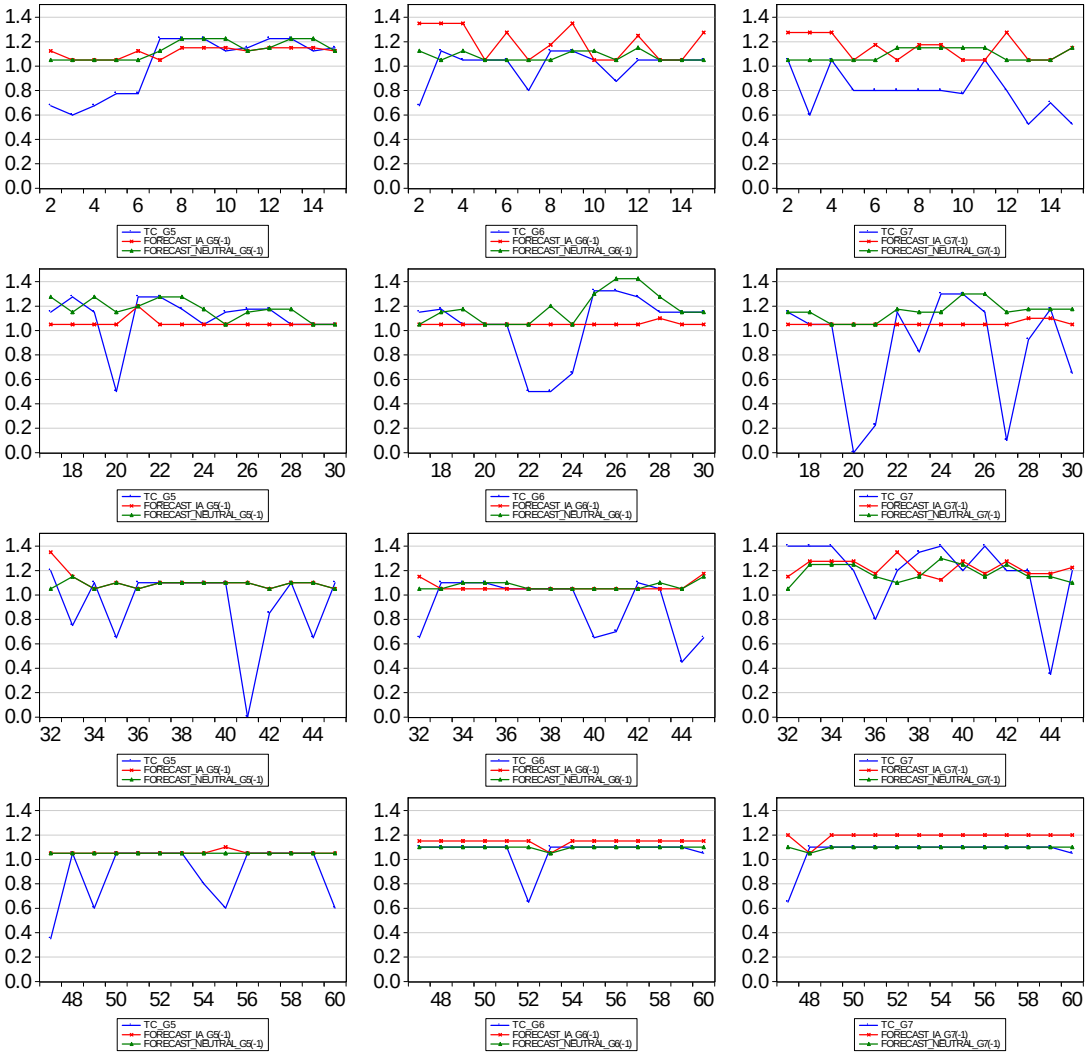


Figure 1.12: The actual total contribution and the predicted total contribution of Groups 5 to 7

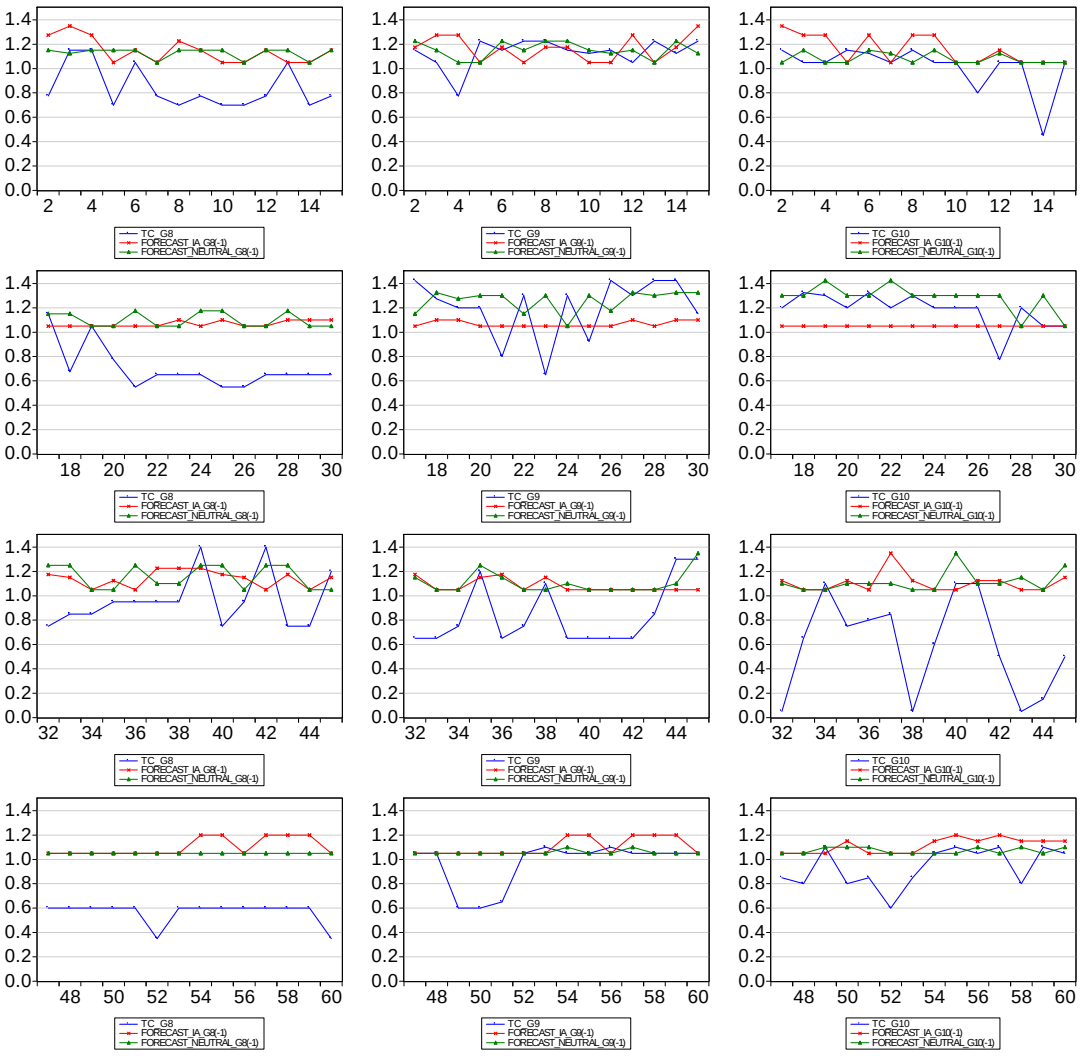


Figure 1.13: The actual total contribution and the predicted total contribution of Groups 8 to 10

preference.

Compared to these neutral predictions, in most cases, the predicted total contributions with inequality-aversion is higher. Both predictions are higher than the profitable threshold during the whole experiment. When subjects have high inequality-aversion, the result is not as unprofitable as we expected. Besides, when the inequality-aversion is not taken into account, the predictions are more stable and closer to the actual outcomes.

When we examine the individual decisions, the predictions with inequality-aversion match the actual decisions by 1838 over 2800 observations (65.6%) while those neutral predictions match the actual decision by 74%. The internal constraint was not supposed to be violated but the results suggest otherwise. In the sample of 1540 observations, the predictions with inequality-aversion match the actual outcome at 77.2% of the observations, while those neutral predictions matched by 84.9%. On the other hand, the predictions on those observations when subjects should follow the external constraint are lower. Amongst the 1260 observations, the predictions with inequality-aversion matched by 51.5% and the neutral predictions match by 61.0%.

To further the discussion, the possible factors are examined by Maximum Likelihood Estimation(MLE) of binary probit regressions. The variables in Table 1.7 are the decision made at the last round (DECISION(-1)), the average number taking Option 1 in the inequality-averse test (INEQ), the year subjects were born (AGE), the political preference from left (1) to right (5) (POLITIC), the religion preference from atheist (1) to religionist (5) (RELIGION), the weakly dominant strategy from not joining (0) to joining (1) (WD STRATEGY), the marginal benefit of the total contribution (γ), and the total

Variable	Probit MLEs(1)	Probit MLEs(2)	Probit MLEs(3)	Probit MLEs(4)	Probit MLEs(5)
Constant term	8.32	0.52***	-9.77	-0.05	11.01
	(12.49)	(0.16)	(20.54)	(0.05)	(16.72)
DECISION (-1)	1.19***		1.36***		1.01***
	(0.07)		(0.13)		(0.09)
INEQ	0.50***	0.81***		-0.15**	
	(0.19)	(0.24)		(0.08)	
AGE	-0.005		0.005		-0.005
	(0.006)		(0.01)		(0.008)
POLITIC	0.05		-0.13**		0.23***
	(0.03)		(0.05)		(0.05)
RELIGION	-0.05**		0.02		-0.17***
	(0.02)		(0.03)		(0.03)
WD STRATEGY	1.16***				
	(0.10)				
γ	-1.27***				-6.45***
	(0.26)				(1.11)
TC (-1)	-0.16		-0.26		-0.36**
	(0.12)		(0.21)		(0.16)
Log Likelihood	-1165.01	-621.21	-515.43	-769.35	-629.48
Total Observation	2520	1500	1400	1120	1120
Observation with decision is 'Join'	1692	1279	1185	507	507

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

Table 1.7: Probit estimations of probability of joining a coalition

contribution of the group at the last round (TC (-1)).

As mentioned earlier, the data of five subjects has been excluded because their attitude to inequality is opposite to our assumption which says the subjects dislike inequality. We examine 45 subjects who have different degrees of inequality-aversion. The estimation of Probit MLEs(1) covers all observations of 2700 decisions which were made individually. Because two variables depend on the outcomes at the last round, only 2520 observations are used for the regression. Amongst these 2520 observations, the subjects decided to join 1692 times and not to join 828 times.

The inequality-averse factor (INEQ), the weakly dominant strategies (WD STRATEGY) and the decision at the last round (DECISION(-1)) have a positive effect on the decision at the 1% significance level. This interesting result implies that the higher inequality-aversion a subject has, the higher incentive this subject has to participate in the coalition. Also, when the decision at the last round or the weakly dominant strategy is being made, the subjects are more likely to choose joining. The marginal benefit of total contribution (γ) has a negative effect on decision-making at the 1% significance level due to the free-riding effect when the subjects' weakly dominant strategy was not to join. Nevertheless, it is insignificant even if the subjects join a coalition in the case where the total contribution at the last round (TC(-1)) is to join. Reviewing the factors listed in the questionnaire, (AGE) and (POLITIC) appear to be statistically insignificant. But, (RELIGION) has a negative effect at the 5% significance level. That means, the more religious a player is, the less likely s/he will join.

It was assumed that the subjects with a higher degree of inequality-aversion were more likely to violate the internal and the external constraints. In or-

der to assess the internal constraint, we use Probit MLE(2) to examine the observations where the subjects' weakly dominant strategy was to join. 85% out of the 1500 observations obeyed the internal constraint. In our hypothesis, the subjects with a higher degree of inequality-aversion were expected to violate the internal constraint, and the coefficient of INEQ should be negative. However, interestingly, the results show that INEQ has a positive effect at the 1% significance level. This striking outcome implies that subjects with a higher degree of inequality-aversion are more likely to to join a coalition. Consequently, this outcome suggests that these subjects with a higher degree of inequality-aversion are less likely to violate the internal constraint. That said, the subjects have stronger incentives to form a profitable coalition when their sense of inequality-aversion is higher. Perhaps due to those subjects's preference of having a fair outcome, a safe act which could keep their payoffs low appears to be more favourable than a risky strategy of punishing other outsiders and forcing them to participate. Those with a lower degree of inequality-aversion tend to act strategicly. They usually attempt to punish free-riders from time to time and force outsiders to participate in a coalition. Such strategic behaviour makes the coalition process unstable over rounds. Comparing the experimental outcomes with the numerical example, we have observed instability in the coalition formation in the experimental results. The experimental results show that the instability is caused by the subjects with low degrees of inequality-aversion rather than those with high degrees.

The estimation of Probit MLE(3) tests the factors included in the questionnaire and the previous results. 1400 observations were collected, except for those in the first round where each sub-treatment was with weakly dominant strategies of joining. The internal constraint was violated 215 times.

The result also supports a significant positive effect on the decision-making at the last round. The effect of (RELIGION) is rather insignificant in this test and (POLITIC) instead has a negative effect at the 5% significance level. It suggests that the pro-right-wingers violating the internal constraint is higher than that of the pro-left-wingers

This result could be explained in the example of Group 9. Four out of five subjects in the group had a switch point in the inequality-averse test. For example, Subject 44 had the highest degree of inequality-aversion - the switch point was at $p = 0.9$. The switch point of subjects 43, 45 and 41 were 0.8, 0.8, and 0.5 respectively. In the membership game, subject 44 violated the internal constraint in only three out of 45 rounds. The violation rates of subjects 43, 45 and 41 are 3%, 0%, and 43%. It shows that the subjects with a higher degree of inequality-aversion were less likely to violate the internal constraint.

However, the internal constraint could be broken by the subjects with a higher degree of inequality-aversion in a few cases. Group 5 where everyone in the group had a switch point in the inequality-averse test as a good example. Subject 21 had the highest degree of inequality-aversion and the switch point is at $p = 0.9$. Following that, the degree of subjects 22 and 24 is $p = 0.8$, the degree of subject 25 is $p = 0.7$, and subject 23 is inequality-neutral - the switch point is at $p = 0.5$. Subject 21 violates the internal constraint in 30% of the 30 rounds, while the violation rates of subjects 22, 24, and 25 are 13%, 0%, and 3% respectively. In this case, the subjects with a higher degree of inequality-aversion are more likely to act against the internal constraint.

The external constraint is assessed by the estimation of Probit MLE(4) where the observations' weakly dominant strategy is not-to-join. The constraint was violated in 45% of the 1120 observations. When a coalition is

unprofitable, it is indifferent whether to join or not. Hence, the subjects would make a random decision in the next round. This is the reason why the external constraint was violated in almost half of the observations.

When a profitable coalition was formed, 44% of the subjects would violate the external constraint in the next round. If we only look at those subjects with a higher degree of inequality-aversion (INEQ$>$ 0.8), only 40% of them violated the constraint. Turning to the results from those with a low degree of inequality-aversion (INEQ$<$ 0.5), the constraint was violated in almost half of the observations. The result shows that the subjects with a high degree of inequality-aversion were more likely to be free-riders. This might appear to be counter-intuitive at first sight, but the subjects with a low degree of inequality-aversion have demonstrated different behaviour of forcing outsiders to participate when their dominant strategy was to join a coalition. When their roles changed to the opposite, they were more likely to compromise and cooperate.

The estimation of Probit MLE(5) examines the factors from the questionnaire. In our hypothesis, the marginal benefit of the total contribution (γ) has a significant negative effect on the decision. In contrast to the experimental evidence of Burger and Kolstad (2010), our results do not support their earlier finding that said that higher marginal benefits would significantly increase a coalition size and consequently the total contribution. This is mainly because our design limits any possible free-riding by excluding the subjects with high marginal benefit. This effect is shown in the estimation of Probit MLE(1). Despite the limitation of our design, the factor of the marginal benefit in the estimation of Probit MLE(5) is significantly negative and corresponds to the earlier findings. Our study provides more detailed information, compared to

the existing literature, about how potential free-riding benefits would weaken the incentives for participation. When the dominant strategy is not to join, higher free-riding benefit comes with higher marginal benefit. The coalition size was likely to be larger than the equilibrium size when the players are with lower marginal benefits.

Our results can be summarised as below

Summary 3

In terms of the coalition formation, the predictions with inequality-aversion does not outperform those without.

In terms of the individual decisions when subjects could free-ride, those with a higher marginal benefit were less likely to join a coalition and prefer to have a lower payoff. On the other hand, the subjects with a high degree of religious belief were more likely to be free-riders by not joining a coalition and having higher payoff.

Right-wingers are more likely to build a larger coalition when they could be free-riders. Comparing to the results on the internal constraint, right-wingers are more likely to violate both internal and external constraints. Right-wingers tend to act strategicly by punishing and compromising when they are in different roles.

1.4 Conclusions

This study has investigated the incentives to participate in IEAs with the other-regarding preferences, particularly the preference of inequality-aversion. The theory used in this study suggests that a stable coalition can be formed both internally and externally, when the signatories have no incentive to leave

and the nonsignatories have no incentive to join. The assumption of inequality-averse preference argues that such a stable coalition would change by considering agents' preferences. Agents with a higher degree of inequality-aversion are more likely to break the internal constraint and leave the coalition.

A two-part experiment has been conducted to validate this theory. The first part was a test to measure the individual attitude to inequality-aversion. The second part was a public good game conducted to mimic the international environmental convention. Subjects were given different payoff tables and asked whether to join or not to join a coalition.

In order to fully capture individual behaviours in an IEA, the experiment has been designed in such a way that teased out as much noise and as many uncertainties as possible. In other words, the theoretical prediction for the experiment was purified to a unique equilibrium. In contrast to the existing literature, the results in this particular design do not support the theoretical prediction that a higher marginal benefit would significantly enlarge a coalition size and the total contribution. On the contrary, the subjects with a lower degree of inequality-aversion are more likely to act strategicly by violating the internal constraint. By doing so, they could force free-riders to participate. But, when their role changes to the opposite, they reacted to compromise their payoffs.

Some other factors inquired in the questionnaire, such as the political preference and religion preference, have also shown significant effects on the decision-making. Pro-right-wingers behave as those with a lower degree of inequality-aversion and make more strategic decisions.

Although it is difficult to generalise solely based on one experiment which has its own limitations in design and data collection, this study has provided

some promising results for understanding the real-world operation of IEAs, especially the dynamics that emerged during the decision making processes. One firm conclusion is that, in order to stabilise a coalition internally, international conventions had better emphasize the importance of fairness to signatories because a high degree of inequality-averse preference would lead a country to participate. An IEA could be enlarged when nonsignatories were informed of the potential damage if the target of the IEA cannot be achieved.

Chapter 2

Altruism in a Climate Coalition

2.1 Introduction

Concerns about potential damages of climate change have grown dramatically over the past decades. Threats and risks emerged from climate change can not be combatted by individual sovereign nation states, actions to reduce greenhouse gas emissions have to be taken at an international level. Several conclusions at the international conventions have been turned into international environmental agreements (IEAs). Well known examples include the Montreal Protocol in 1987 and the Kyoto Protocol in 1997.

A huge number of literature has explored the structures of and variations of IEAs (Barrett, 1994 and 2001; Bahn *et al.*, 2009; Weikard *et al.*, 2006 and Bratberg *et al.*, 2005). Typically, these studies are based on the assumption that agents pursue their self-interest, thereby the models used to investigate IEAs are based on individual countries' explicit welfare and ignore the effects of externalities. Results from these studies also show that the number of signatories in an IEA decreases when the benefit of global abatement increases.

However, a growing number of experimental evidences has challenged such rational self-interest (Willinger and Ziegelmeyer, 2001 and Kolstad, 2014). Altruistic behaviours and high degrees of cooperation are rather common in experimental observations on public-good provision (Fischbacher *et al.*, 2001). For example, the theoretical work of Grüning and Peters (2010) suggests that countries's abatements and level of participation in an IEA increase when the countries's preferences incorporate justice and fairness. Hence, social (other-regarding) preferences have become a non-neglectable factor in the studies of IEAs.

In the previous chapter, inequality-averse preferences have been introduced by examining the effect of the diversity of individual payoffs. The results in Chapter 1 show that the degree of individual inequality-aversion is an important variable to the decisions in an IEA membership game. Individuals care not only about their own payoffs but also the gaps between theirs and other's payoffs. The diversity of individual payoffs is a negative factor when agents would like to approach a fair outcome.

In this chapter, I provide another approach of modelling other-regarding preferences, which are agents' altruistic behaviours in this case. Nagel (1970) defines altruism as 'not abject self-sacrifice, but merely a willingness to act in the consideration of the interests of other persons, without the need of ulterior motives' (1970, p. 79). Unlike the concept of inequality-aversion, altruistic agents care about the overall welfare of all agents rather than the variance of individual payoffs. On the one hand, if fairness is the only goal of IEAs, the result of a minimal variance of individual payoffs may be meaningless if everyone abates nothing and no IEA is formed. On the other hand, altruistic agents might have a stronger incentive for participating in IEAs, if agents

would like to maximise the overall welfare by cooperating in the coalition.

The importance of altruistic preferences has been recognised in recent studies of IEAs. Van der Pol *et al.* (2012) consider altruism in the participation decision of a two-stage IEA game. Two types of altruism are studied in their paper: impartial altruism, where countries show a concern for all other countries, and community altruism, where the concern is extended only to coalition partners. They claim that certain degree of altruism is sufficient to stabilise a grand coalition. On the other hand, Hahn and Ritz (2014) relax the assumption so that altruistic preferences may not reflect directly on player's behaviour on the membership status. Their model allows strategic behaviours that a player could behave different to her true preference. In this model, they propose a hypothesis that a country almost always behaves less altruistically than its true preference. Hahn and Ritz claim that it may be difficult to infer social preferences from this observed behaviour.

Both arguments of van der Pol *et al.* (*ibid*) and Hahn and Ritz (*ibid*) have not been examined with empirical evidences. It is the goal of this study to examine their model with experimental evidences and to provide a different explanation.

Having said that, the aim of this chapter is to explore the effects of altruistic preferences on individual incentives of participating in an IEA. Specifically, it investigates how altruism may help individuals to overcome free-riding and join an IEA. For this purpose, a model is built with agents who have different levels of concerns about the overall payoff for all countries. Their individual altruistic preferences affect their decisions about whether they would like to join an IEA or not. Having said that, this chapter does not explore the difference of individual payoffs, but the sum of coalition payoffs.

The examination of the impacts of altruistic preferences is based on a novel experimental design, which comprises two parts. The first part is a test to examine individuals' degrees of altruistic preferences. Subjects are asked to answer a series of give-or-take questions. Their altruistic preferences are indicated by how many times the subjects give away rewards to a stranger. The second part of the experiment is a repeated one-shot public good game. Each subject has different marginal benefits of the total contribution to a public good. This particular design provides better observations on individual behaviours than the previous design of identical marginal benefits did.

The chapter is structured as follows. In Section 2, a model based on the assumption of altruistic preferences is presented. A numerical example is provided to illustrate the impact of a high degree of altruistic preferences on the coalition formation. In Section 3, an experiment with two parts is described in detail to test the theory. The instructions are included in Appendix 1.4. Section 4 shows the experimental outcomes and the data analyses. The conclusions are in the final section.

2.2 The model

The framework is that of N heterogeneous countries, indexed $k = 1, ..., N$. A country k's welfare is

$$\pi_k = (-x_k) + \gamma_k X$$

where the individual abatement x_k is standardised between 0 (pollute) and 1 (abate), and $\gamma_k \in [0, 1]$ is country k's individual marginal benefit of the global

abatement X, while $X = \sum_{k=1}^{N} x_k$.

Supposed that n countries ($n \in [2, N]$) decide to form a coalition. We assume that countries are heterogeneous with respect to various marginal benefits of the total abatement. We rank their marginal benefits from high to low as $\gamma_1 > \gamma_2 > ... > \gamma_N$. Since the main interest of this study is to examine the motivations for participation in an IEA, we simplify the situation to a one-stage membership game by assuming that signatories would abate and nonsignatories would pollute. If a profitable coalition is formed, the members in the coalition abate to maximise their joint payoff. Their aggregate benefit of the total abatement is larger than the cost ($\sum_{k=1}^{n} \gamma_k > 1$). Nonsignatories would pollute while receiving a free-riding benefit from the coalition. On the other hand, if the coalition is unprofitable, all countries would pollute and have nothing for return.

The coalition payoff $\prod$ is the sum of all signatories' pre-redistribution payoffs as

$$\begin{aligned} \prod &= \sum_{i=1}^{n} \pi_i \\ &= \sum_{i=1}^{n} [\gamma_i n - 1] \end{aligned}$$

where γ_i is signatory i's marginal benefit of the total abatement.

As mentioned in Chapter 1, the coalition members using transfers to equalise net payoffs between agents may be an inferior assumption in studying IEAs. This mechanism suggests a less unequal distribution of payoffs through transferring. Under this assumption, the countries with higher marginal benefit of the total abatement are more likely to leave the coalition, because those coun-

tries could earn higher payoff for the absence. Because the coalition members make a common decision and share the responsibility of maximising the coalition payoff, it is adequate to assume the coalition payoff is equally shared by the members.

Hence, the post-redistribution payoff of a signatory i can be presented as

$$\pi_s = \sum_{i=1}^{n} \gamma_i - 1 \tag{2.1}$$

Since members in the coalition cooperate to abate, the payoff is the aggregate payoff of signatories net of the cost of abatement. If the coalition is profitable, the payoff is positive.

The payoff of a nonsignatory j is

$$\pi_j = n\gamma_j \tag{2.2}$$

where γ_j is a nonsignatory j's marginal benefit of the total abatement. Because nonsignatories do not pay for abatement, each of them can enjoy the free-riding benefit, which is the size of the coalition times its own marginal benefit.

Following Hahn and Ritz (2014), we build an altruism objective function of a country k

$$\begin{aligned} S_k &= (1-\theta_k)\,\pi_k + \theta_k W \\ &= \pi_k + \theta_k \sum_{k' \neq k} \pi_{k'} \end{aligned} \tag{2.3}$$

where $\theta_k \in [0,1]$is country k's degree of altruistic preference, π_k is country k welfare while $W = \sum_{k=1}^{N} \pi_k$ is the global welfare. It is intuitive to assume

$\frac{\partial S_k}{\partial \theta_k} > 0$, i.e. k's welfare is positively correlated with the magnitude of altruistic preference. Besides, $\frac{\partial S_k}{\partial \pi_{k'}} \geq 0$ means that the higher the payoff of the other is, the higher the welfare country k has. The function can be presented as the self-interest payoff of country k and its altruism concern about the aggregate payoff of other k'(all countries except k) countries. In other words, the objective of this chapter is the sum of payoffs rather than the variance of payoffs.

The problem of the nonsignatory j is as follows:

The individual welfare of a signatory i is its own payoff and the adjusted payoffs from other countries. Hence, the maximising problem of a signatory i is as follows:

$$\begin{aligned} \max S_i &= \begin{cases} \pi_s + \theta_i \sum_{s' \neq s} \pi_{s'} & \text{if } \sum_{i=1}^{n} \gamma_i \geq 1 \\ 0 & \text{if } \sum_{i=1}^{n} \gamma_i < 1 \end{cases} \\ &= \begin{cases} (\sum_{i=1}^{n} \gamma_i - 1) + \theta_i \left[\sum_{i' \neq i}^{n-1} (\sum_{i=1}^{n} \gamma_i - 1) + \sum_{j}^{N-n} \left(n\gamma_j \right) \right] & \text{if } \sum_{i=1}^{n} \gamma_i \geq 1 \\ 0 & \text{if } \sum_{i=1}^{n} \gamma_i < 1 \end{cases} \end{aligned} \tag{2.4}$$

On the other hand, the welfare of a nonsignatory j depends on its own payoff and the adjusted payoffs from others. The problem of the nonsignatory j is therefore as follows:

$$\begin{aligned} \max S_j &= \begin{cases} \pi_j + \theta_j \sum_{j' \neq j} \pi_{j'} & \text{if } \sum_{i=1}^{n} \gamma_i \geq 1 \\ 0 & \text{if } \sum_{i=1}^{n} \gamma_i < 1 \end{cases} \\ &= \begin{cases} \left(n\gamma_j \right) + \theta_j \left[\sum_{i=1}^{n} (\sum_{i=1}^{n} \gamma_i - 1) + \sum_{j}^{N-n-1} \left(n\gamma_j \right) \right] & \text{if } \sum_{i=1}^{n} \gamma_i \geq 1 \\ 0 & \text{if } \sum_{i=1}^{n} \gamma_i < 1 \end{cases} \end{aligned} \tag{2.5}$$

Given that n^* is the smallest profitable coalition, if a signatory i decides to join an IEA, the country follows the *internal constraint* as

$$S_i^s(n^*) \geq S_i^{ns}(n^* - 1) \tag{2.6}$$

Similar to the explanation in Chapter 1, the left-hand-side of the inequality (2.6) is i's welfare when it is a signatory in a n^*-member coalition. The right-hand-side of the inequality is i's welfare if it decides to be a nonsignatory and the size of coalition becomes $(n^* - 1)$. Because the externality of abatement, everyone is benefited by the abatement of a single country. When a country decides to leave an IEA, all countries have to suffer its absence with a decreasing abatement level. The gap between the objective of being a signatory and that of being a nonsignatory is enlarged with a higher degree of altruistic attitude. Therefore, the internal constraint becomes more robust when others' payoffs are taken into account of the objective function.

If a nonsignatory j decides not to join an IEA, the country follows the *external constraint* as

$$S_j^{ns}(n^*) \geq S_j^s(n^* + 1) \tag{2.7}$$

The left-hand-side of the inequality (2.7) means j's welfare when it is a nonsignatory with an n^*-member coalition. The right-hand-side of the inequality is j's welfare if it decides to become the $(n^* + 1)$-th member in the coalition. Since the benefit of abatement of being a signatory is enlarged with a higher degree of altruistic attitude, the external constraint could be violated.

Given that all agents are self-interested, when the internal, the external constraints and the unique equilibrium condition $\left(1 + \gamma_{n^*} > \sum_{i=1}^{N} \gamma_i\right)$ are all

satisfied, we have learnt from Chapter 1 that there is a unique equilibrium coalition. However, when the agents are with varying altruistic attitudes, the unique equilibrium may not exist.

Conjecture 4

Depending on the individual degree of altruism, the size of a coalition could be n^* or larger than n^*.

This conjecture is based on the possible outcomes shown in Appendix 2.1.

The internal constraint is always satisfied no matter to agent's attitude to altruism. The stronger attitude to altruism an agent has, the less likely the agent would violate the internal constraint. This is due to the coalition is designed to enhance the overall payoffs. The utility of an altruist in a coalition is higher than that of an egoist in the same coalition. On the other hand, the external constraint could be violated if a nonsignatory has a high degree of altruistic attitude. To sum up, the coalition size could be enlarged if a subject has strong attitude to altruism. This conjecture will be tested by the following experiment. When individual altruistic preferences are measured, their individual decisions in the membership game and the coalition formation could be predicted by this conjecture.

2.3 Experiment design

The experiment is incorporated into the game designed for Chapter 1. As mentioned in the previous chapter, the instructions (see the Appendix 1.3) were provided on the subjects' desks. The instructions consisted of three parts. Since the purpose of this study was to investigate the impacts of altruistic

preferences on the coalition formation, the data from part 2 and part 3 in the instructions could satisfy our targeting.

Two-part design was used in this experiment. The first part (Part 2 in the instructions) provides the indicator of individual altruistic preferences. The second part (Part 3 in the instructions) was a membership game in which subjects were asked whether or not to join a coalition. Subject were given different payoffs for their decisions. The experiment in detail is illustrated as follows.

Altruism test

The design of the altruism test renovates Bettinger and Slonim (2006)'s and Andreoni and Miller (2002)'s experiments. In this test, subjects were paired but without knowing each other beforehand. Each subject answered a series of give-it-or-take-it decisions in 20 rounds. Their payoffs were affected by their own decisions as well as their partners'. In order to get unbiased data, the subjects did not know the decisions made by their partners.

Each subject was given 1 token as an endowment. He or she (the *dictator*) decided where the token would go to himself/herself or another subject (called *receiver*, a random subject in the lab). All subjects were playing the role of dictators. Though they were also receivers to their opponents, the payoffs as receivers were only released in the end of the experiment. The token for the dictator was denoted as T_1 [1], and the number for the receiver was denoted as $(1 - T_1)$. The decision was made by the dictator, and the receiver could only accept. The value of the token was different to the dictator and the receiver

[1] The token is indivisible, hence T_1 is either 0 or 1.

(z_1 and z_2 respectively). Hence, the payoff of the dictator was $T_1 z_1$ and the receiver's was $(1 - T_1) z_2$.

The welfare function of the dictator was

$$\begin{aligned} S_1 &= T_1 z_1 + \theta_1 (1 - T_1) z_2 \\ &= T_1 (z_1 - \theta_1 z_2) + \theta_1 z_2 \end{aligned}$$

The dictator could keep the token when the level of altruism was as small as $\theta_1 < \frac{z_1}{z_2}$, otherwise the token would go to the receiver. The exchange rate ($\frac{z_1}{z_2}$) was descending over rounds. The highest level of altruism was assumed as 1 and altruistic neutrality was assumed as 0. The altruism level could be found with the decreasing exchange rate over rounds. We can find the approach value by asking the subjects with 20 various sets of exchange rates. The possible payoffs set for subjects in the experiment is listed in Table 2.1. The payoffs in the left column are what a dictator had when she or he decided to *keep* the token (Option 1). The payoffs in the right column are what a receiver had when the dictator decided to *give* the token (Option 2).

The expected decision in the first round was to keep the token (Option 1). Hence, $T_1 = 1$ implies that the weight of individual's own payoff is higher than that of other agents' payoffs. When the same question repeats over rounds, depending on individual altruistic attitudes, each agent would change their minds from taking to giving the token at a particular round. After this round, agents sacrifice their own payoffs to benefit others without being able to ask for any reward. Thus the altruism level θ can be inferred.

Round	z_1	z_2	$(\frac{z_1}{z_2})$
1	£1	£1	1
2	£10	£10.5	0.95
3	£7.5	£8	0.94
4	£5	£5.5	0.91
5	£2.5	£3	0.83
6	£7.5	£10	0.75
7	£5	£7.5	0.67
8	£0.5	£1	0.5
9	£5	£10.5	0.48
10	£2.5	£5.5	0.46
11	£1	£2.5	0.4
12	£2.5	£7.5	0.33
13	£2.5	£10	0.25
14	£0.5	£2.5	0.2
15	£1	£5.5	0.18
16	£1	£7.5	0.13
17	£0.5	£5	0.1
18	£1	£10.5	0.095
19	£0.5	£7.5	0.07
20	£0.5	£10	0.05

Table 2.1: List of values of the token and exchange rate

Experiment of a coalition game

This part is a joint experiment with Chapter 1 and the design of the experiment has been illustrated in detail in the previous chapter. The summary of the coalition game is as follows. The subjects were randomly assigned to groups of five subjects. They did not know who they are playing with, but they knew that they were playing with the same people during the whole session. Tables were provided with the individual payoffs of all possible coalition combinations. They did not know others' decisions until everyone made their decisions. The history of the membership status and payoffs of all subjects in the group were revealed on their screen at the end of each round.

The results are reported as follows.

2.4 Experimental results and analyses

In the altruism test, a selfish and rational subject would always decide to *take* (Option 1) for 20 rounds. On the other hand, an altruistic subject would decide to *give* (Option 2) at round 20 for sure. The higher the degree of altruistic preference a subject has, the more likely the subject would choose Option 2. Since the ratio of exchange rate ($\frac{z_1}{z_2}$) has been ranked in the order from high to low, this order indicates the level of altruistic preference at the switching point where an altruistic subject alters her or his decision from Option 1 to Option 2. The table in Appendix 2.2 shows the result of the altruism test of all subjects.

Figure 2.1 shows the effect of altruism. It is perhaps unsurprising that all subjects decided to keep the token in the first round. However, the smaller

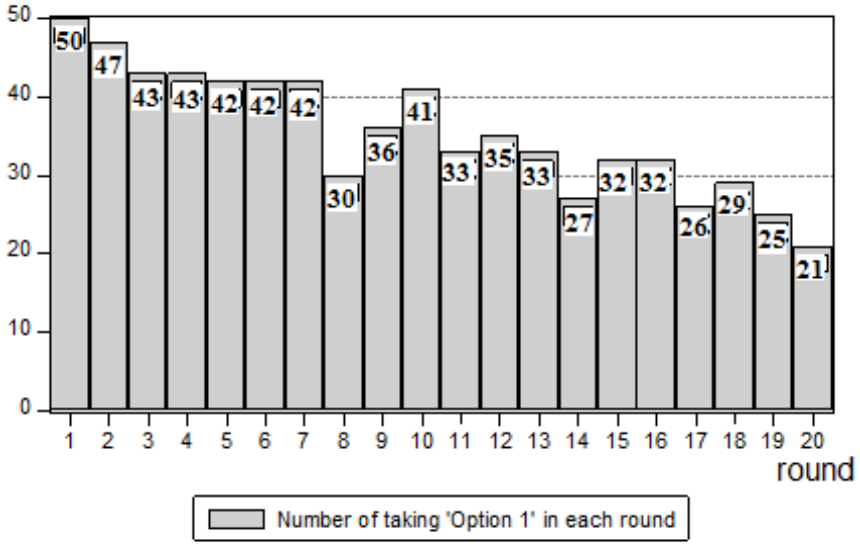


Figure 2.1: Number of subjects taking 'Option 1' in each round

the ratio of the exchange rate was, the more likely the subjects gave up the token. In the last round, almost 60% of subjects gave up £0.5 and made an unknown partner earning £10. The remaining 18 subjects could be considered as egoists, because they chose Option 1 throughout.

The majority of the subjects had altruistic preferences, as we observed them giving up their allowances to benefit unknown partners. 13 of them had consistent behaviour with one switching point from Option 1 to Option 2. In general, a decreasing trend in Figure 2.1 shows that the degree of the subjects' altruistic preferences is heterogeneous.

In addition, it is also interesting that the value of the token is an important factor to the subjects' decision-making. When the opportunity cost of giving was £0.5 in rounds 8, 14, and 17, the subjects were more likely to behave altruistically. Compared to the results in the next rounds (rounds 9, 15, and 18), the number of taking 'Option 1' was lower even when the ratio of exchange

Variable	Altruism level OLS Regression		
Constant term	−9.17		
	(20.17)		
AGE	0.005		
	(0.01)		
POLITIC	0.03		
	(0.05)		
RELIGION	−0.06*		
	(0.03)		
Log Likelihood	−10.4767	R-squared	0.07
Total Observation	50		

Note: Each cell contains coefficient and standard error in parenthesis
* means 10% significant level

Table 2.2: OLS estimation of altruistic preference

rate was higher in rounds 8, 14, and 17.

Table 2.2 shows the OLS estimation of altruistic preference. The dependent variable is the average times of taking the Option 1 in the inequality-averse test. Independent variables are the subject's' age (AGE), political attitude (POLITIC), and religious attitude (RELIGION). At a 10% significance level, the impact of religious attitude is negative. This interesting result implies that the subjects who identified themselves with stronger religious belief behaved less altruistically. Later, the factor of religious attitude also has a significant effect on the membership decisions. This striking result contradicts our intuition that many religious believers are volunteers doing charity work. The rest factors have insignificant impacts on the subjects' altruistic attitudes.

The results on the coalition game have been reported in Chapter 1. There is no need to repeat them again here. This chapter analyses the impact of altruistic preferences on the coalition formation with two methods. Firstly, the

predicted coalition formation with altruistic preferences is generated. Comparing to the prediction with a self-interested preference, the predictions with altruistic preferences are more closed to the actual total contributions. Secondly, the factors which may influence the subjects' decisions on a public good game are examined by Maximum Likelihood Estimation(MLE) of binary probit regressions. In addition, a comparison of the results with those of inequality-aversion assumption in Chapter 1 helps our understanding of the impacts of different social preferences.

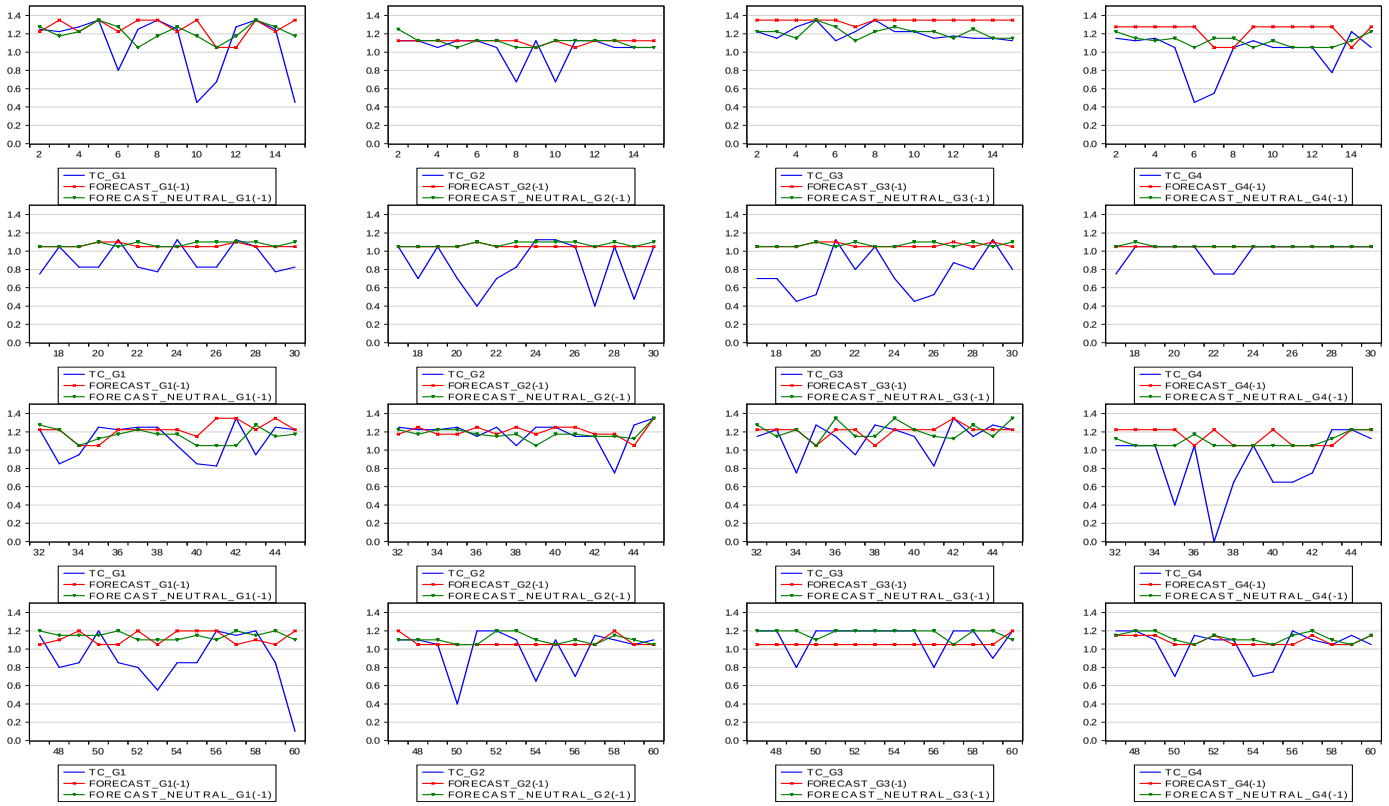


Figure 2.2: Actual total contribution and predicted total contribution with altruism of Groups 1 to 4

By using the data of individual altruistic preferences in the altruism test and the historical records of the decisions in the coalition-game experiment, a predicted coalition formation with the altruistic preferences is generated. To benchmark this prediction, a self-interested prediction is generated with the historical records of the decisions in the coalition game only. Figure 2.2 presents the total contribution of Groups 1 to 4. The blue solid line in each chart presents the actual total contribution in a sub-treatment. Given the historical outcomes in the past round and the subjects' individual altruistic preferences, the altruistic predictions of the coalition formation are generated as the red short dashed line with cross. In order to examine the precision and robustness of our model for measuring the effect of altruistic attitude, the predictions which are generated with historical data only are shown as the green dash line with triangle marks. The predictions are called the self-interested predictions. The actual total contribution and the predicted total contribution of Groups 5 to 10 are shown in Figures 2.3 and 2.4.

Both predictions are higher than the actual total contribution in general. Compared to the self-interested predictions, the predictions which consider the altruistic attitude is closer to the actual total contribution in most cases. Moreover, the variance of the predictions with altruistic attitude is usually higher than that of the self-interested predictions.

When we examine the individual decisions, the predictions with individual altruistic attitudes match the actual decisions by 2074 over 2800 observations (74.1%). The predictions are slightly better than the self-interested predictions which match the actual decision by 73.6%. When the subject's dominant strategy is to join the coalition, over the 1540 observations, the predictions with altruistic attitude match the actual decisions by 84.9%. The performance

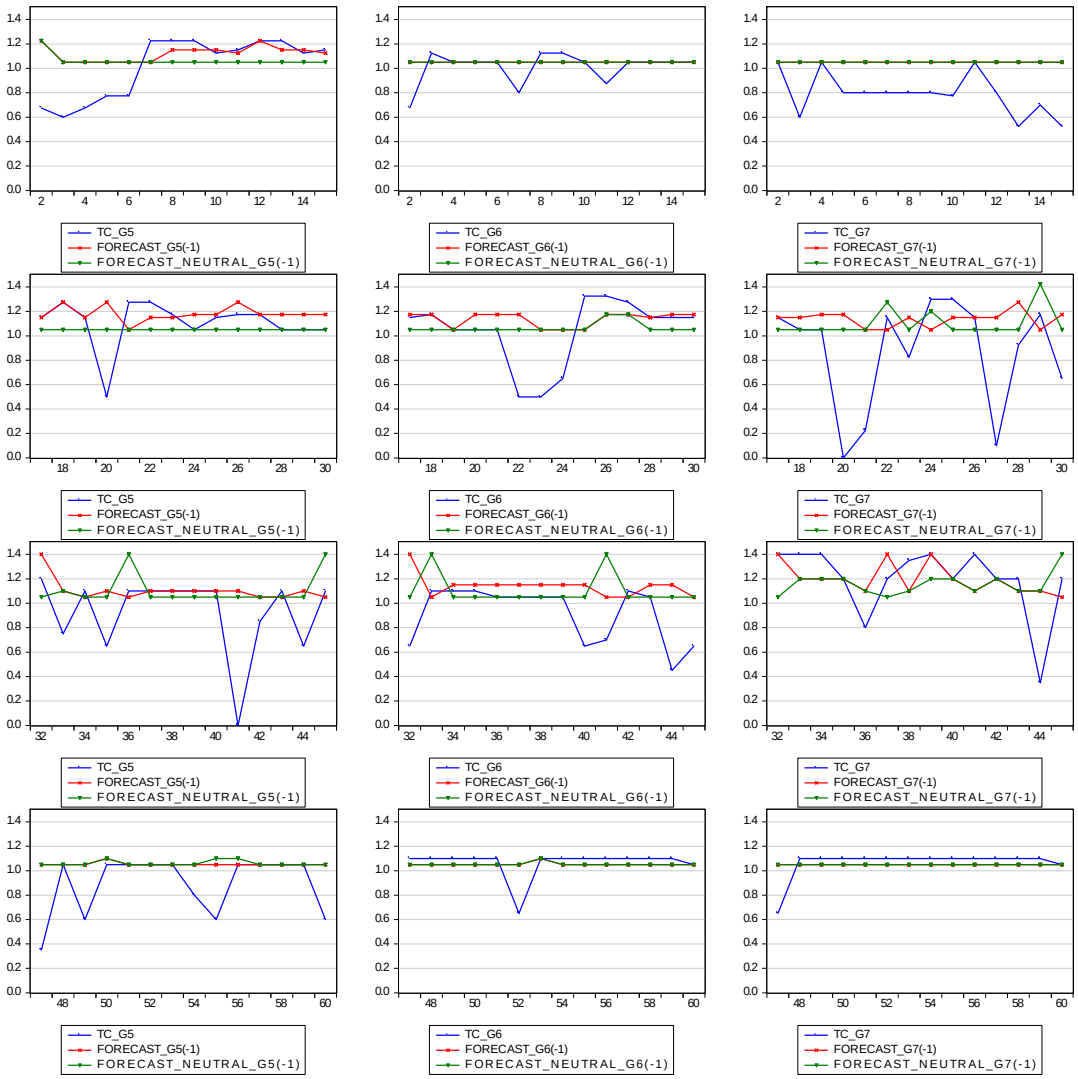


Figure 2.3: Actual total contribution and predicted total contribution with and without altruistic preferences of Groups 5 to 7

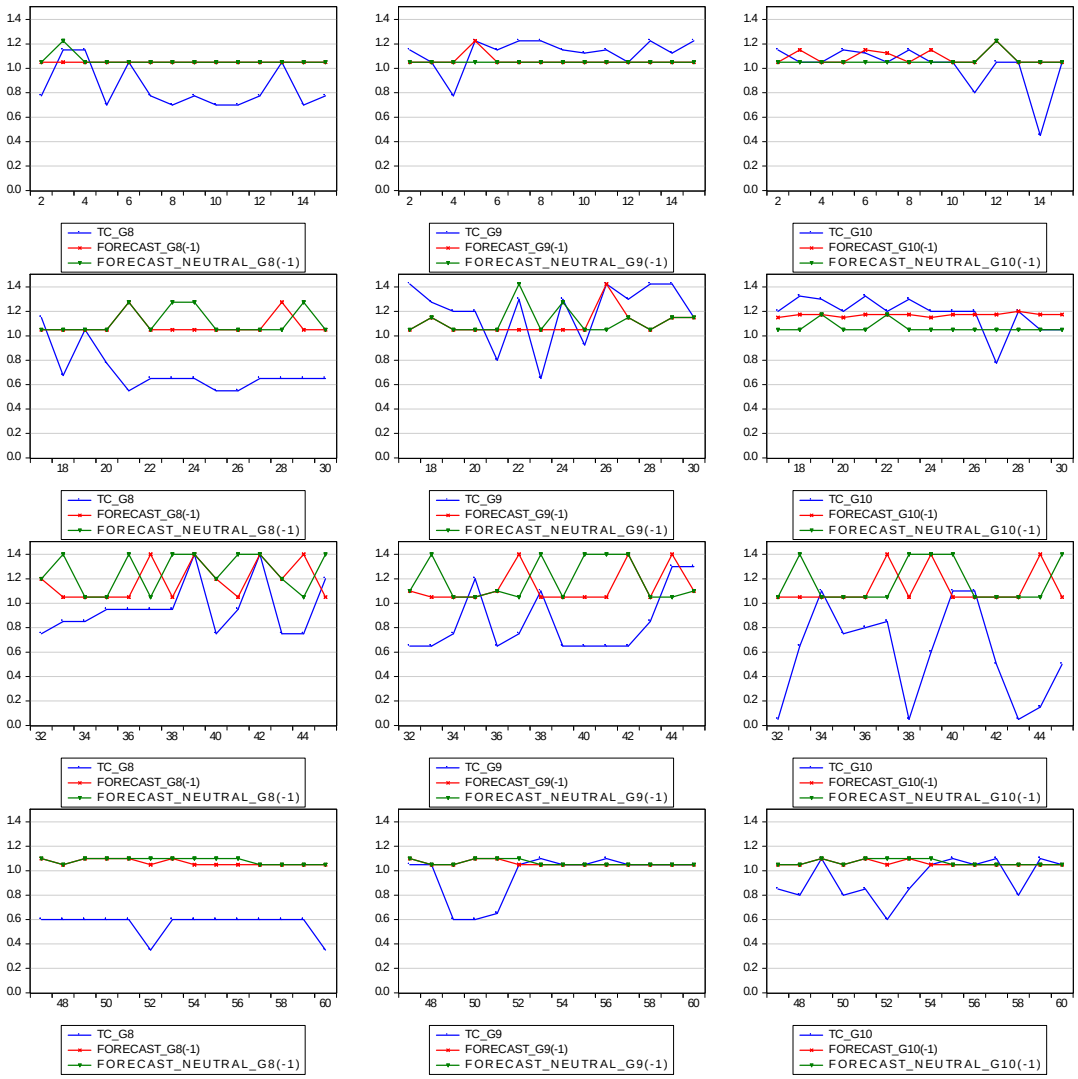


Figure 2.4: Actual total contribution and predicted total contribution with and without altruistic preferences of Groups 8 to 10

of the self-interested predictions is at the same matching rate. On the other hand, when the subject's dominant strategy is not to join the coalition, over the 1260 observations, the predictions with altruistic attitude match by 60.8% and the self-interested predictions had slightly lower rate by 59.7%.

In order to investigate the impact of altruistic preferences on the individual decisions, Maximum Likelihood Estimation(MLE) of binary probit estimations are reported in Table 2.3. The variables are the decision made at the last round (DECISION(-1)), the average number choosing 'take' in the altruism test (ALTRU), the year subjects were born (AGE), the political preference from left (1) to right (5) (POLITIC), the religion preference from atheist (1) to religionist (5) (RELIGION), the weakly dominant strategy from not joining (0) to joining (1) (WD STRATEGY), the parameter of marginal benefit of total contribution (γ) and the total contribution of the group at the last round (TC (-1)).

The regression of Probit MLE(1) employs all observations with a total of 3000 individual decisions from the coalition game. Since two variables (DECISION (-1) and TC (-1)) depend on the outcomes at the last round, 2800 observations are used for the regression. The subjects decide to join in 1884 times and not to join in 916 times. The decision at the past round and the weakly dominant strategy have a positive effect on the decision at the 1% significance level. It means when joining is either the decision made in the past round or the subject's weakly dominant strategy, the subject is more likely to join. Meanwhile, the amount of the total contribution at the past round (TC(-1)) has an insignificant effect. For the factors from the questionnaire, the factors of AGE and POLITIC are statistically insignificant while the factor

Variable	Probit MLE(1)	Probit MLE(2)	Probit MLE(3)	Probit MLE(4)	Probit MLE(5)
Constant term	-4.22	0.98***	-9.01	-0.27***	0.34
	(11.74)	(0.09)	(20.12)	(0.09)	(17.30)
DECISION (-1)	1.12***		1.28***		0.98***
	(0.07)		(0.12)		(0.08)
ALTRU	0.13	0.10	0.16	0.25**	0.01
	(0.09)	(0.12)	(0.14)	(0.11)	(0.13)
AGE	0.002		0.005		-0.0002
	(0.006)		(0.01)		(0.009)
POLITIC	0.04		-0.11**		0.16***
	(0.03)		(0.05)		(0.04)
RELIGION	-0.04*		0.01		-0.10***
	(0.02)		(0.03)		(0.03)
WD STRATEGY	1.14***				
	(0.09)				
γ	-1.12***				-5.21***
	(0.24)				(1.05)
TC (-1)	-0.07		-0.19		-0.14
	(0.11)		(0.20)		(0.14)
Log Likelihood	-1321.26	-683.96	-568.88	-930.17	-737.33
Total Observation	2800	1650	1540	1350	1260
Observation with Membership=1	1884	1410	1308	629	576

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

Table 2.3: Probit estimations of probability of joining a coalition

of RELIGION has negative impact. It implies that the stronger religious belief a subject has, the less likely to join the IEA. As we mentioned earlier, the religious attitude also influences the subjects' altruistic preferences. This will lead to the factor of altruistic attitude has insignificant effect on the membership decision.

The data can be divided into two groups: a group of observations where subjects' dominant strategy is joining the coalition and another group where subjects' dominant strategy is not to join. In other words, the internal constraint is examined by the regression of Probit MLE(2). There are 1650 observations which are with the weakly dominant strategy to join the coalition. The internal constraint is not satisfied 240 times. The result does not show a significant impact of altruistic preferences on individual decisions.

On the other hand, the external constraint is examined by the regression of Probit MLE(4). There are 1350 observations which are with the weakly dominant strategy not to join the coalition. The external constraint is not satisfied in 629 times. Interestingly, at the 5% significance level, the lower the degree of altruistic preference a subject has, the more likely the subject is to violate the external constraint and participate in the coalition. The subjects' behaviour is in contrast to the self-interested prediction as well as our intuition.

This could be explained in the example of group 5. Four of five subjects in the group have a switching point in the altruism test. Subject 24 has the highest degree of altruism since this subject has given away the token from an early round. Subjects 21, 23, and 25 have changed their minds at rounds 4, 17 and 19 respectively. Subject 22 has two switch points at rounds 5 and 8. In the membership game, subject 24 with a high degree of altruism only violates the external constraint in 1 of 15 rounds, while subject 23 who has low

altruistic attitude violates the external constraint at 47% of 30 observations.

However, the altruistic prediction happens in some cases. For example, subject 25 with a low degree of altruism violates the external constraint in only 3 of the 30 rounds. Subjects 21 and 22 have a high degree of altruism and violate the constraint in 63% and 57% of 30 rounds, respectively.

Having said that, even the subjects with similar degrees of altruism behave differently in the coalition game. For example, the subjects in groups 8 and 9 keep the token throughout 20 rounds. It implies that they all have a very low degree of altruism. However, in the coalition game, the subjects in group 8 form a profitable coalition in 12 of 60 rounds but the subjects in group 9 form a profitable coalition in 42 over 60 rounds. Subjects 38, 42 and 43 are not bound with the external constraint over 50% of the rounds where they are better not to join the coalition.

The regressions of Probit MLE(3) and Probit MLE(5) assess other factors which might influence subjects' decisions. The factor of the decisions in the past round is positive at 1% significant level. It means their preferences are rather consistent.

The marginal benefit of total contribution (γ) in the regression of Probit MLE (5) has the negative effect on the decision at the 1% significance level. It it intuitive that the higher the free-riding benefit a subject has, the less likely it is that the subject would like to contribute to a public good. Other variables, the total contribution in the past round and the age of subjects, are insignificant factors.

Subjects are playing a more complicated strategy to cooperate with each other. Compared to the outcomes which exclude the observations in the first round, subjects are more likely to cooperate at the first round in each sub-

treatment when they do not know the decisions of each other. In 200 observations collected from the first round of each sub-treatment, the internal constraint held at more than 90% of 110 observations while the external constraint is violated in almost 60% (53 out of 90 observations). On the other hand, such high cooperation rates decrease after seeing other subjects' decisions. The internal constraint is satisfied in 85% of 1540 observations and the external constraint is violated in 46% of 1260 observations.

Other interesting variables include the political preferences. When subjects are better off to cooperate, pro-right-wing supporters are more likely to break the internal constraint by not joining the coalition. However, when the subjects have the chance to free ride, pro-right-wingers are more likely to give up this chance. It seems that right-wingers are more strategic by using punishing and cooperating to increase the overall welfare. In addition, it is interesting that the subjects who consider themselves to have a high degree of religious belief are less likely to give up the chance of free-riding.

2.4.1 Comparison of results for altruistic and inequality-averse preferences

Compared to the results in Chapter 1, the model with altruistic preferences performs better than that with inequality-averse preferences. In terms of the individual social preferences and the membership status in the past round, the predictions on the individual membership decisions are generated. When the subjects' weakly dominant strategy is to join the coalition, over the 1540 observations, the predictions with altruistic preferences match 84.9% of the actual decisions while the predictions with inequality-averse preferences match

77.2%. Meanwhile, the predictions with no other-regarding preferences perform as good as the predictions with altruistic preferences.

On the other hand, when the subjects' weakly dominant strategy is not to join the coalition, over the 1260 observations, the predictions with altruistic preferences matches 60.8% of the actual decisions while the the predictions with inequality-averse preferences match by 51.5%. Overall, the performance of the predictions with altruistic attitude is superior than those with inequality-averse preferences.

In terms of results of the probit regressions, the degree of altruistic preference is an significant negative factor to the probability of joining a coalition. Especially when subjects have the chance to free-ride, the lower the altruistic preferences a subject has, the more likely it is that the subject would cooperate. Similarly, the impact of the degree of inequality-averse preference is different to our expectation. The higher degree of inequality-aversion a subject has, the subject is more likely to cooperate where the expected decision is not to join the coalition. It seems that the social preferences may not show directly as the behaviour in an interactive game.

The experimental evidences in both Chapter 1 and Chapter 2 show the strategic behaviour in the public good game. Hahn and Ritz (2014) also claim that it may be difficult to infer countries' true preferences for altruism from their observed behaviour. The challenge to the further studies and policy makers on climate negotiations is to find out the linkage between the preferences and the behaviour.

2.5 Conclusions

This chapter examines the impact of altruistic preferences on the formation of IEAs. Existing experimental literature (such as Fischbacher et al., 2001) suggests that subjects often behave altruistically in a public good game. To test this, a particular model is built with individual altruistic attitudes. The theoretical result shows that agents who have a higher degree of altruistic preferences are more likely to cooperate. If an agent with a high degree of altruistic preference plays, there may exist a larger coalition than the Nash prediction.

In order to examine the model, a two-part experiment was designed and run. In the first part, the altruism test questioned if subjects would give away their benefits to an unknown partner or not. Their altruistic preferences were indicated with the number of rounds, in which they sacrifice without any reward. Following this, subjects are asked to play a public good game.

The data on individual altruistic preferences provides valuable information that about half of the subjects have different degrees of altruistic preferences. Two type of predictions are generated: the first one uses the historical records of the individual decisions in the membership game and the individual altruistic attitudes; the second type uses the historical records only. Both type of predictions are higher than the actual total contribution in general. Compare to the neutral predictions, the predictions which consider the altruistic attitude is closer to the actual total contribution in the most cases. Moreover, the variance of the predictions with altruistic attitude is usually higher than that of the neutral predictions.

Compared to the actual individual decisions, the predictions with indi-

vidual altruistic attitudes have better performance than the self-interested predictions. When subject's dominant strategy is to join the coalition, the predictions with altruistic attitude perform same to the self-interested predictions. On the other hand, when subject's dominant strategy is not to join the coalition, the predictions with altruistic attitude are slightly superior to the self-interested predictions.

The estimations illustrate the subjects' motivations. The rate of cooperation in a coalition game seems to be negatively correlated with the magnitude of altruistic preferences: the lower the degree of altruistic preference, the more is the cooperation. This is particularly so when the subjects' weakly dominant strategies are not to join a coalition.

Chapter 3

Sustainability and International Environmental Agreements

3.1 Introduction

This chapter examines the relation between perceptions of sustainability and the formation of international environmental agreements (IEAs) in a cross-generational model with a two-stage game in two periods.

Human activities have left many enduring footprints and legacies. As a result, the ecosystems on the Earth have changed dramatically due to the rapid industrial development in the past decades. Our society is now facing a range of environmental crises. Actions are urged to maintain basic needs of the future generations, because the outcome of human development is often irreversible and will be passed on to the next generations. Some of the environmental problems can be addressed at the national level. As an effective supra-national governmental authority that can handle cross-border environmental issues has not yet existed, IEAs have served as the second-best solution.

The most common purpose of the existing IEAs is to assure sustainable development. The term 'sustainable development' was first used in the report of *Our Common Future* which was published by the World Commission on Environment and Development (WCED) in 1987. In that publication, it is defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs".

Lately, 'sustainability' or 'sustainable development' have become buzzwords overloaded with fuzzy meanings. At the discussion of IEAs, stakeholders such as governments, industries, NGOs, trade unions, academics all have different understandings of 'sustainability'. For instance, the objective of the United Nations Framework Convention on Climate Change (UNFCCC) in 1992 declared "... Such a level should be achieved within a time-frame sufficient to allow ecosystems to adapt naturally to climate change, to ensure that food production is not threatened and to enable economic development to proceed in a sustainable manner (UNFCCC, 1992, p. 4)". Later in 1997, the UNFCCC stated in the Kyoto Protocol that "Each Party included in Annex I, in achieving its quantified emissions limitation and reduction commitments under Article 3, in order to promote sustainable development (Kyoto Protocol, 1997, Article 2)".

The report of *Our Common Future* links sustainability with poverty eradication, equitable distribution of benefits derived from natural resources, population policies, development of human activities and maintenance of natural resources. Although efforts have been endeavoured to construct a common standard between different international bodies, and negotiation has been undergoing to reduce the distance between the representatives and the represented, there still exists an epistemic gap in various perceptions and interpre-

tations of 'sustainability'. This is especially the case when one considers the tension between future generations and democracy (social diversity or different cultures, intrinsic values or resources for local problems, mute actors, multiple representations, different issues, may they be techno-centred, eco-centred, anthropo-centred). All these factors shape individual decision-makings and contribute to dynamics, and stability of an IEA.

To explore how these different understandings of 'sustainability' shape individual decisions and incentives to join (or not to join) an IEA, this paper will focus on individual concerns about the future generations.

Based on a literature review, the concept of sustainability can be categorised at three levels: individual, societal, and the ecosystem levels.

To individuals, sustainability usually means to achieve constant utility (Solow, 1974 and Hartwich, 1977) and avoid any decline in utility (Pearce et al. 1989; Pezzey, 1997). More precisely, Pezzey (*ibid*) identifies three distinct constraints: sustainable level, sustained level, and survivable level. Here, utility is the objective for individuals to achieve sustainability.

To society, sustainability is when the basic needs of the future generations are satisfied (WCED, 1987); the length of the existence of the human race is maximised (Georgescu-Roegen, 1971); the present value of the social welfare is not declining (Riley, 1980); and the per capita incomes of the future generations are no worse off (Pearce *et al*, 1989). The indicators of sustainability at this level are the theoretical social welfare and the practical figures (such as Green Net National Product expanded by Hartwick, 1977 and Genuine Savings provided by Hamilton and Clemens, 1999).

Moving to the ecosystem, sustainability covers a wide range of objectives which include exhaustible natural resources (Meadows *et al*, 1972), renewable

natural resources, production waste, and biological diversity. In order to meet sustainability, exhaustible resources, such as minerals and fossil fuel deposits, have to be extracted at a rate at which the length of use is maximised. Renewable resources, such as fisheries and forests, have to be harvested at a natural and manageable speed of regeneration. In addition, biological diversity also has to be maintained for the basic need of the survival development.

The previous studies have proposed three types of policy goals for sustainability: (1) achieving constant or non-declining individual utility function (Solow, 1974 and Pezzy, 1997); (2) avoiding any decline in social values from the present time onwards (Riley, 1980); and (3) maintaining existing 'safe minimum standards' (Toman, 1994). These can be applied onto management of natural exhaustible resources and renewable resources and waste emissions (Solow, 1974 and Stiglitz, 1974).

In order to avoid any decline in social present value, Woodward (2000) identifies a set of behaviours that would lead to sustainable life; these behaviours entail intergenerational fairness. This means that the future generations will not envy the present one, and there exists an alternative, feasible choice that there is no envy between generations. Woodward's ethical assumption emphasises the current generation's responsibility to future generations. That said, the current generation has to consider not only their present welfare but also the welfare of future generations. Woodward's concept of sustainability emphasises the fairness across generations.

Toman (1994) discusses the concept of 'safe minimum standard' when speaking of strong sustainability. Because human activities in natural environments have 'irreversible' effect, the human capital can not substitute the natural assets when decision makers have low level of information but high

potential asymmetry in the payoff. Hence, Barbier and Markandya (1990) impose a minimum stock of environmental assets. In this model, when the asset is driven below this safety criterion, environmental degradation will destroy the natural clean-up and regenerative processes in the environment. Following this concept, Martinet (2011) proposes an approach that defines the objectives of sustainability using sustainability threshold indicators.

Though the concept of sustainability is so important to IEAs, relatively few attention has been paid to discuss the relationship between this key factor and the formation of IEAs. The majority of theoretical studies employs static models to analyse the coalition formation (e.g. Barrett, 1994, 2005; Yi, 1997 and Carraro and Siniscalco, 1998). These models ignore the importance of sustainability and simply assume that humans are immortal because there exists a static optimal pollution level where humans welfare is maximised.

However, these static models do not reflect the reality to illustrate the impact of sustainability which emphasises the fairness between generations. The present generation thinks and behaves differently from future generations, even though they might care the future generation. Recent studies (e.g. Germain *et al.* 2003; de Zeeuw, 2008; Rubio and Ulph, 2007) have employed some more dynamic models to describe human development in the infinite horizon. However, this setting still presumes that future generations are always richer than the present generation in terms of welfare, hence exclude the possibility of decreasing welfare. That said, the cross-generational fairness is hardly considered in the literature. To our best knowledge, this study is the first to consider the impact of sustainability (more specifically the impact of diverse perceptions of sustainability) in the formation of IEAs. In order to model the impacts of different perceptions of sustainability, the value of the social welfare of the

future generation has to be taken into account when reviewing the present generation's welfare and decision-making. Additionally, the non-declining social welfare also needs to be reconsidered. The sustainability criterion dictates that the social welfare of the future generation should not be worse than that of the present generation.

This chapter builds a two-stage game in two periods. In each period, the decision makers are different agents. They decide whether or not to participate in an IEA in the first stage. In terms of their membership status, countries will decide the emissions level in the second stage. We consider two scenarios for the objective function in Period 1. To examine the effect of different perceptions of sustainability on the formation of IEAs, a myopic (MYO) scenario is first proposed. In the MYO scenario, the decision makers of the old generation care about their own welfare. Following, we consider the model in the sustainable development (SD) scenario that the decision makers of the old generation care about that of the young generation. The old generation attempts to maximise the over-generational welfare and ensure that the welfare of the young generation is no worse off than the young one.

Our result shows that the marginal cost of the total emissions plays an important role. The higher the marginal cost is, the lower the individual emissions level. A grand coalition formation is possibly formed when the marginal cost is very small. Besides, the awareness of sustainability have small but ambiguous impact on the formation in two periods.

The chapter is structured as follows. In Section two, a two-stage game in two periods model is built in two scenarios. A numerical example presented in Section three illustrate the coalition formation in different scenarios. The conclusion and discussion are in the final section.

3.2 The model

Unlike Chapters 1 and 2, this chapter investigates the cross-generational preferences based on a model that focuses on the frameworks of IEAs and ignore individualities. This assumption of identical countries is drawn on Barrett (1994), Rubio and Ulph (2007) and Breton *et al.* (2010) which assume countries are homogeneous in their analyses of incentives of participating in IEAs. We appreciate to the assumption of heterogeneous players, however, we have emphasised the point in the introduction: to our best understanding, there is no paper which model sustainability in the discussion of the formation of IEAs.

In order to investigate the long term effect of pollution, we present a model of a two-stage game in two periods. Table 3.1 shows the decision process of the model. The decision makers live for one period only: the old generation lives in Period 1 and the young generation lives in Period 2. In each period, there is a two-stage game: in the first stage membership game, the countries decide whether or not to participate in an IEA. In the second stage emission game, countries make the decision on the level of emissions in terms of their membership status. Nonsignatories choose emissions in a non-cooperative way to maximise their own payoff, while signatories act as one to maximise the coalition payoff. The emission plan is irreversible, but the total stock of emissions will accumulate with a certain decay rate. Hence, the total stock of emissions is the sum of the accumulated emissions from the past and the aggregated emissions in that period. In order to understand the importance of sustainability in IEAs, the focus of this study is on the coalition formation in two scenarios. The young generation have the same objective function in both

Time horizon	Period 1	Period 2
Player	Old generation	Young generation
2-stage game	Membership game | Emission game	Membership game | Emission game
Total emission	$E_1 = \delta E_0 + \sum_{i=1}^{n_1} e_{i.1} + \sum_{j=n_1+1}^{N} e_{j.1}$	$E_2 = \delta E_1 + \sum_{i=1}^{n_2} e_{i.2} + \sum_{j=n_2+1}^{N} e_{j.2}$
Objective function (MYO scenario)	Nonsignatory : $\pi_{j,1}$ Signatory : Π_1	Nonsignatory : $\pi_{j,2}$ Signatory : Π_2
Objective function (SD scenario)	Nonsignatory : $\pi_{j,1} + \beta\pi_{j,2}^f$ s.t. $\pi_{j,1} \leq \pi_{j,2}^f$ Signatory : $\Pi_1 + \beta\Pi_2^f$ s.t. $\Pi_1 \leq \Pi_2^f$	Nonsignatory : $\pi_{j,2}$ Signatory : Π_2

Table 3.1: The decision process of the model

scenario, however, the old generation have different objective functions. While countries concern about only the welfare of the old generation in the MYO scenario, countries in the SD scenario concern about not only the welfare of the old generation but also that of the young generation. In addition, the welfare of the young generation is required to be no worse than that of the old generation.

There is a finite set of N identical countries. While there obviously are other capital stock variables in abatement (e.g. non-fossil power stations), we only consider the stock of pollutant in the model. The pollutant is a by-product of production, the stock of pollutant has a strong positive correlation with industrial processes. The normalised benefit function from the production

can be presented as

$$B\left(e_{k,t}\right) = \frac{1}{b} e_{k,t}^{b}$$

where $e_{k,t}$ denotes a country k in Period t has to choose a level of emissions, $k \in \{1, ..., N\}$ and $t \in \{1, 2\}$ [1]. The parameter b is the benefit elasticity of emission where $b \in (0, 1)$. This assumption of a concave benefit function implies the diminishing rate of returns. It says that as additional units of emissions are generated, eventually the marginal benefit from the production will decrease. It should be noted that the benefit elasticity of emission b is a constant and determined by available technology level, or management of the production process. Higher benefit elasticity by advanced technology implies a country has a higher benefit per unit of emissions. This elasticity measures the responsiveness of benefit to a change in level of emissions stock. For example, when $b = 0.5$, a 1% increase in emissions stock would lead to approximately 0.5% increase in benefit.

While the pollutant also causes severe damage to the environment, the cost for country k is highly correlated with the global stock of emissions. The damage cost function for k is a linear function denoted as

$$C\left(E_t\right) = \gamma E_t$$

where γ is the marginal cost of the total stock of emissions E_t where $\gamma > 0$. The total stock of emissions contains the accumulated emissions from the past and

[1] Each country chooses a level of emissions for the production, we do not have a particularly reason to normalise the level to 1.

the aggregate emissions generated by the signatories and the nonsignatories

$$E_t \equiv \delta E_{t-1} + \sum_{i=1}^{n} e_{i.t} + \sum_{j=n+1}^{N} e_{j.t} \tag{3.1}$$

Suppose n of N countries[2] join an IEA and the rest are nonsignatories. We define $e_{i,t} \geq 0$, $i = 1, ..., n$ and $t = 1, 2$, as the individual emissions of a signatory i in Period t. By controlling an equal amount of emissions in each signatory, the optimal coalition payoff can be reached. On the other hand, $e_{j,t}$, $j = n+1, ..., N$ and $t = 1, 2$, denotes the individual emissions of a nonsignatory j in Period t.

Hence, (3.1) can be read as the total stock of emissions is the sum of the accumulated emissions from the past, the emissions from signatories and the emissions from nonsignatories in the current period. The accumulated emissions from the past depends on the natural decay factor per period $\delta \in (0, 1)$. Because Greenhouse gas (GHG) stock is absorbed naturally over time, the total pollution decays over time. It is reasonable to assume that the decay rate is between zero and one. Because the stock of emissions is accumulative, the decision on emissions which is generated by the old generation affects to not only the old generation but also the young generation.

We assume that all countries decide their emissions plan simultaneously. In period t, a country k's net benefit function is

$$\pi_{k,t} = B\left(e_{k,t}\right) - C\left(E_t\right)$$

Each generation lives for one period and optimises the welfare with respect

[2] n is an integer value between 0 and N.

to its current level of emissions as

$$\max_{e_{k,t}} \pi_{k,t} = \left[\frac{1}{b} e^{b}_{k,t} - \gamma E_t\right] \tag{3.2}$$

As mentioned previously, given the initial stock of the pollutant, there is a two-stage games:

- In the first stage, countries decide whether or not to join an IEA.
- In the second stage, countries decide their emission in terms of their membership status.
  - Signatories move as one by determining a common emissions level to maximise the coalition welfare.
  - Nonsignatories decide their own emissions level to maximise their own individual welfare.

When we discuss the formation of self-enforcing IEAs, following Rubio and Ulph (2007), the membership of any country is determined by a random process such that the probability of any country being a signatory in that period is simply the membership of the stable IEA in that period divided by the total number of countries. This probability is the same for all countries, but the membership of countries in different periods could be different. Two scenarios in the decision process have been shown in Table (3.1): (i) myopic (MYO), (ii) sustainable development (SD). The young generation faces the same objective function, while the old generation have different policy goals in both scenarios. In the MYO scenario, the old generation is myopic and the decision makers only concern their own welfare in Period 1. In the SD scenario, the old generation

concerns not only its own welfare but also the expected welfare of the young generation. Besides, the sustainability criterion dictates that the welfare of the young generation cannot be worse than the welfare of the old generation.

We would like to highlight that for the SD scenario the expected welfare of the young generation is based on the membership status of the old generation. In Period 1, the old decision makers have the expectation and belief about the membership of the young generation when they consider the cross-generational welfare. This assumption is adequate because practical IEAs do not usually have an expire date[3]. The young generation is expected to inherit the membership from the old generation. However, in Period 2, the membership status of the young generation does not necessary be the same to that of the old one. In other words, the coalition formation could be different in both periods[4].

We solve the two-stage and two-period game by backward induction. Section 3.2.1 discusses the young generation's decisions on the two-stage game which includes the emission plan and the membership status in Period 2. Then we discuss the old generation's decisions on the two-stage game in Period 1. There are two scenarios: Section 3.2.2 discusses the myopic scenario where the old generation cares about its welfare; Section 3.2.3 discusses the sustainable development scenario where the old generation cares not only its welfare but also the young generation's.

[3]For example, the Montreal Protocol on Substances that Deplete the Ozone Layer in 1987.

[4]For example, the Kyoto Protocol has two commitments periods. The first commitment period applies to emissions between 2008-2012, and the second commitment period applies to emissions between 2013-2020. Only 37 parties have stated to participate in the second commitment period. Others (e.g. Belarus, Kazakhstan and Ukraine) may withdraw from the Protocol or not put into legal force the Amendment with second round targets.

3.2.1 Decisions in Period 2

Second-stage emissions game

Regardless of the decision makers are myopic or not, the young generation faces the same decision process. Suppose that n_2 countries has decided to participate in the coalition in Period 2, so that the rest $(N - n_2)$ countries are nonsignatories. From (3.2), a young nonsignatory j maximises its objective function that its individual payoffs

$$\max_{e_{j,2}} \pi_{j,2} = \left[\frac{1}{b} e_{j,2}^{b} - \gamma E_2 \right] \tag{3.3}$$

where $e_{j,2}$ is the emissions level of a nonsignatory j in Period 2. The total emissions $E_2 = \delta E_1 + \sum_{i=1}^{n_2} e_{i.2} + \sum_{j=n_2+1}^{N} e_{j.2}$ is the sum of the accumulated stock of emissions in the past period with the decay rate δ and the aggregated emissions from signatories and nonsignatories in Period 2.

The optimal level of emissions of a young nonsignatory is

$$e_{j,2} = (\gamma)^{\frac{-1}{1-b}} \tag{3.4}$$

Since the parameter b is set between 0 and 1, we therefore learn that a higher marginal cost of the total emissions ($\gamma > 1$) leads to a lower optimal emissions level ($\frac{\partial e_{j,2}}{\partial \gamma} < 0$). The derivative with respect to the parameter b of the emissions level ($\partial e_{j,2}/\partial b$) is ambiguous [5]. When the marginal cost of

[5] A simple proof is below:
(1) take logarithms of both side $\ln(e) = \frac{-1}{1-b} \ln(\gamma)$
(2) take the derivative with respect to b, $\frac{\partial \ln(e)}{\partial e} \frac{\partial e}{\partial b} = \frac{-\ln(\gamma)}{(1-b)^2}$
So $\frac{\partial e}{\partial b}$ is positive when γ is less than 1 and it is negative when γ is greater than 1 .

total emissions is smaller than 1, it implies that the higher technology level may incur more pollution. In light of the history of human development, the more advanced technology we have, the more we would like to produce. While the technologies are more efficient and produce fewer pollutants per unit of product, the level of emissions increases due to the increasing consumption of products. In other words, the advanced level of technology development may lead to a more efficient production per unit of emission, but it also encourage countries to emit more in total. On the other hand, when the marginal cost of total emissions γ is greater than 1, the more advanced technology would lower the emissions level. This is due to when the marginal cost is high, the increase on the pollution cost is faster than the growth of benefit by the technology development.

The emissions level of a signatory i is determined when the coalition payoff is optimised with regard to the common emissions level $e_{i,2}$, $\forall i \in 1, ..., n_2$

$$\max_{e_{i,2}} \Pi_2 = \sum_{i}^{n} \left[\frac{1}{b} e_{i,2}^{b} - \gamma E_2 \right] \tag{3.5}$$

All signatories make a common decision to maximise the coalition payoff. If the number of the coalition is n_2, the coalition emissions is n_2 times of a signatory i's emissions level. It is presented as $\frac{\partial E_t}{\partial e_{i,t}} = n_2$. This group effect implies that having more signatories brings a stronger influence on the global emissions quantity.

Therefore, the optimal emissions level of a young signatory i in Period 2 is

$$e_{i,2} = (n_2 \gamma)^{\frac{-1}{1-b}} \tag{3.6}$$

$\frac{\partial e_{i,2}}{\partial n_2} < 0$ and $\frac{\partial e_{i,2}}{\partial \gamma} < 0$ mean the size of the IEA and the marginal cost of the total emissions are negative to the optimal emissions level of a signatory. It implies that larger a coalition is, lower each member country emits. This is due to the group effect, where the larger group make a higher impact on the total emission. In other words, a higher total abatement level could be made by a larger coalition. Also, the high marginal cost would lead to a low emissions level. However, the technology parameter (b) has ambiguous effect on the emissions level.

The payoffs of countries in two periods are

$$\pi_{j,2} = \frac{1}{b}(\gamma)^{\frac{-b}{1-b}} - \gamma\left[\delta E_1 + n_2(n_2\gamma)^{\frac{-1}{1-b}} + (N-n_2)(\gamma)^{\frac{-1}{1-b}}\right] \quad (3.7)$$

$$\pi_{i,2} = \frac{1}{b}(n_2\gamma)^{\frac{-b}{1-b}} - \gamma\left[\delta E_1 + n_2(n_2\gamma)^{\frac{-1}{1-b}} + (N-n_2)(\gamma)^{\frac{-1}{1-b}}\right] \quad (3.8)$$

All individuals will be benefited when the coalition is enlarged ($\frac{\partial \pi_{j,2}}{\partial n_2} > 0$ and $\frac{\partial \pi_{i,2}}{\partial n_2} > 0$). We also learnt that a nonsignatory j has a higher benefit than a signatory i and everyone pays the same cost, hence the welfare of a nonsignatory is higher than that of a signatory.

First-stage membership game

In order to find the formation of an IEA, we follow D'Aspremont *et al.* (1983), a n_2^*-member stable coalition exists when two constraints are satisfied

$$\pi_{j,2}(n_2^* - 1) \leq \pi_{i,2}(n_2^*) \quad (3.9)$$

$$\pi_{i,2}(n_2^* + 1) \leq \pi_{j,2}(n_2^*) \quad (3.10)$$

Here, $\pi_{i,2}$ is the payoff when an old country decides to participate in an

IEA and $\pi_{j,2}$ is the payoff of the country who decides not to participate. The number in the parenthesis means the size of the IEA. The *internal constraint* (3.9) implies the incentive of participation of a signatory i. A country would participate in a coalition as one of n_2^* member countries only if being a signatory is better than being a nonsignatory. When the constraint is not satisfied, that country would withdraw from the coalition. When the number of signatories decreases and the coalition is no longer profitable, the consequence is that the IEA could no longer exist. On the other hand, the *external constraint* (3.10) explains the incentive of a nonsignatory. A country would stay away from a coalition when the payoff of being a nonsignatory is better than that of being the $(n_2^* + 1)$-th member. When both constraints are satisfied, the coalition is considered as *stable*.

Following, Section 3.2.2 and 3.2.3 discuss the decisions of the old generation in Period 1 in two scenarios. The decision process of the two-stage game is: countries firstly decide whether or not to participate in an IEA, then decide their emissions plan in relation to their membership status. The game is also solved by backward induction.

3.2.2 Decisions in Period 1 in the Myopic (MYO) scenario

Second-stage emissions game

In the myopic scenario, the decision makers care about the welfare in Period 1 only. Similar to the objective function of the young generation, suppose there are n_1 members in the IEA in Period 1, an old nonsignatory j maximises only

its payoff with respect to its individual emissions level ($e_{j,1}$)

$$\max_{e_{j,1}} \pi_{j,1} = \left[\frac{1}{b} e_{j,1}^{b} - \gamma E_1\right] \tag{3.11}$$

where $e_{j,1}$ is the emissions level of a nonsignatory j in Period 1, and the total stock of emissions E_1.

Hence, the optimal emissions level of j is obtained from (3.11). The myopic old generation emits the same level as the young generation does.

$$e_{j,1} = (\gamma)^{\frac{-1}{1-b}}$$

On the other hand, the coalition attempts to maximise the aggregate payoff in Period 1 with respect to the common emissions level $e_{i,1}$

$$\max_{e_{i,1}} \Pi_1 = \sum_{i}^{n} \left(\frac{1}{b} e_{i,1}^{b} - \gamma E_1\right) \tag{3.12}$$

The optimal emissions level of a myopic old signatory i is at the same to that of a young signatory.

$$e_{i,1} = (n_1\gamma)^{\frac{-1}{1-b}}$$

The post-distribution payoffs of a myopic signatory i and a myopic nonsignatory j in period 1 are

$$\pi_{j,1} = \frac{1}{b}(\gamma)^{\frac{-b}{1-b}} - \gamma\left[\delta E_0 + n_1 (n_1\gamma)^{\frac{-1}{1-b}} + (N - n_1)(\gamma)^{\frac{-1}{1-b}}\right] \tag{3.13}$$

$$\pi_{i,1} = \frac{1}{b}(n_1\gamma)^{\frac{-b}{1-b}} - \gamma\left[\delta E_0 + n_1 (n_1\gamma)^{\frac{-1}{1-b}} + (N - n_1)(\gamma)^{\frac{-1}{1-b}}\right] \tag{3.14}$$

First-stage membership game

The stable coalition in Period 1 can be found with the two constraints by D'Aspremont *et al.* (1983)

$$\pi_{j,1}\left(n_1^* - 1\right) \leq \pi_{i,1}\left(n_1^*\right) \tag{3.15}$$

$$\pi_{i,1}\left(n_1^* + 1\right) \leq \pi_{j,1}\left(n_1^*\right) \tag{3.16}$$

Here, $\pi_{i,1}$ is the post-redistribution payoff when a country decides to participate in an IEA and $\pi_{j,1}$ is the payoff of that country decides not to participate. The number in the parenthesis means the size of the IEA in Period 1.

The *internal constraint* (3.15) implies the participation incentive of a signatory i. A country would participate in a coalition as one of n_1^* member countries only when being a signatory is better than being a nonsignatory. When the constraint is not satisfied, that country would withdraw from the coalition. When the number of signatories decreases and the coalition is no longer profitable, the consequence is the IEA would collapse. On the other hand, the *external constraint* (3.16) explains the incentive of a nonsignatory. A country would stay away from a coalition when the payoff of being a nonsignatory is better than that of being the $(n_1^* + 1)$-th member. When both constraints are satisfied, the coalition is considered as *stable*.

It should be noted that IEAs being formed in the beginning of each period, the coalition formation in Period 1 (n_1) does not necessary remain until Period 2 (n_2). The emissions level and the welfare will be affected by the number of

signatories, both the emissions level and the welfare could be different in both periods. Given that that the coalition size remains the same for two periods ($n_1 = n_2$), a young and an old generation emit at the same level. From (3.1), we have learnt that the young generation has to suffer an extra cost from the accumulated emissions. The young generation's welfare is worse than the old generation's. According to the concepts of sustainability, this can be labelled an unsustainable system.

The outcome of the myopic scenario is summarised as follows.

Summary 5 *In the myopic scenario, nonsignatories generate the same level of emissions in two periods.*

Suppose that the coalition size remains the same for two periods, the system is unsustainable where the optimal levels of emissions for the two generations are the same but the welfare of the young generation is worse than that of the old one.

3.2.3 Decisions in Period 1 in the Sustainable development (SD) scenario

The result from the MYO scenario shows that myopic decision makers would generate the same level of emissions in two periods. Their welfare of two generation depend on the parameters of the benefit and cost functions, as well as the coalition formation in each period. In order to ensure a sustainable system, we now restructure the model for the sustainable development (SD) scenario in Period 1. The social welfare of the young generation would be no worse than that of the old generation. The two-stage game is also solved by backward induction.

Second-stage emissions game

In the SD scenario, the old generation considers not only the welfare in Period 1 but also that of that in Period 2. Let $\pi^f_{j,2}$ denote the *expected* welfare of the young generation under the coalition formation in Period 1. In practise, IEAs do not usually have an expiry date. When the old generation make the decision in Period 1, it is reasonable to assume that the young generation inherits the membership from the old generation. The expected coalition formation in Period 2 remains the same to the formation in Period 1. Given that there are n_1 signatories to an IEA in Period 1, the expected number of signatory in Period 2 would be n_1 [6]. In terms of its membership status in Period 1, the old generation predicts the emissions level and the welfare of the young generation.

An old nonsignatory j's objective function is

$$\max_{a_{j,1}} \pi_{j,1} + \beta \pi^f_{j,2} = \left(\frac{1}{b} e^b_{j,1} - \gamma E_1\right) + \beta \left(\frac{1}{b} e^b_{j,2} - \gamma E_2\right) \tag{3.17}$$

$$\pi_{j,1} \leqq \pi^f_{j,2} \tag{3.18}$$

where β is the discount factor attached by one generation to the welfare of the next [7]. Given the goal of sustainability which is to maximise the cross-generational welfare, a nonsignatory j cares not only about the payoff at present but also the payoff in the future. It implies an intergenerational altruism, which means that the current generation does not ask for anything in return from the future generation. The higher value of β, higher is the weight

[6] However, the young generation reforms the coalition and decides its actual membership in Period 2. The young generation does not have to follow the expectation of the old generation.

[7] The discount factor β is assumed in the range of 0 and 1. It implies the weight of how much the old generation cares about the young generation.

put on the young generation by the old generation.

Inequality (3.18) refers to the sustainability criterion of which the welfare of the future generation is no worse than that of the present generation[8]. It implies intergenerational fairness which denotes that the present generation does not sacrifice the future welfare. When the payoff of the old generation is higher than that of the young generation, the constraint is bounded and the old generation will adjust the emissions level to maintain the intergenerational fairness.

We therefore set up the Lagrange function with respect to $e_{j,1}$ as

$$\mathcal{L}_j\left(e_{j,1}\right) = \pi_{j,1} + \beta\pi_{j,2}^f + \lambda_j\left(\pi_{j,2}^f - \pi_{j,1}\right) \tag{3.19}$$

The Kuhn-Tucker conditions for the maximisation problem in (3.19) are

$$\frac{\partial \mathcal{L}_j}{\partial e_{j,1}} = -\gamma\left[\left(1+\beta\gamma\right) - \lambda_j\left(1-\delta\right)\right] + \left(1+\lambda_j\right)e_{j,1}^{b-1} = 0,\ e_{j,1} \geq 0 \tag{3.20}$$

$$\frac{\partial \mathcal{L}_j}{\partial \lambda_j} = \pi_{j,2}^f - \pi_{j,1} \geq 0,\ \lambda_j \geq 0,\ \lambda_j\left(\pi_{j,2}^f - \pi_{j,1}\right) = 0 \tag{3.21}$$

The members in the coalition will attempt to maximise the coalition payoff over periods. The expected payoff Π_2^f is under the same membership status.

[8]We appreciate that it is unusual to impose a non-declining welfare criterion in a two period model where welfare in Period 1 is compared with welfare just in Period 2. Given that pollutant in Period 2 is unaffected by what happens in Period 1, to ensure that welfare in Period 2 exceeds welfare in Period 1, it will be necessary to reduce welfare in Period 1 significantly. This may not be a very satisfactory model with which to study the impact of the non-declining welfare constraint. However, the constraint is adequate to the concepts of sustainable development.

The objective function of the old generation is

$$\begin{aligned} \max_{e_{i,1}} \Pi_1 + \beta \Pi_2^f &= \sum_i^{n_1} \left(\frac{1}{b} e_{i,1}^b - \gamma E_1 \right) + \beta \sum_i^{n_1} \left(\frac{1}{b} e_{i,2}^b - \gamma E_2 \right) \quad (3.22) \\ \Pi_1 &\leqq \Pi_2^f \quad (3.23) \end{aligned}$$

This can be rewritten in a Lagrangian with respect to $e_{i,t}$ as

$$\mathcal{L}_i \left(e_{i,t} \right) = \Pi_1 + \beta \Pi_2^f + \lambda_i \left[\Pi_2^f - \Pi_1 \right] \quad (3.24)$$

The Kuhn-Tucker conditions for the maximisation problem in (3.24) are

$$\begin{aligned} \frac{\partial \mathcal{L}_i}{\partial e_{i,1}} &= -\gamma n_1 \left[(1 + \beta\gamma) - \lambda_i \left(1 - \delta \right) \right] + \left(1 + \lambda_i \right) e_{i,1}^{b-1} = 0, \ e_{i,1} \geq 0 \quad (3.25) \\ \frac{\partial \mathcal{L}_i}{\partial \lambda_i} &= \Pi_2^f - \Pi_1 \geq 0, \ \lambda_i \geq 0, \ \lambda_i \left[\Pi_2^f - \Pi_1 \right] = 0 \quad (3.26) \end{aligned}$$

To solve the problem, we discuss in the following cases:

Case 1. No criterion is binding ($\lambda_j = \lambda_i = 0$)

When no criterion is binding, $\lambda_j = \lambda_i = 0$. From (3.20) and (3.25), we yield the optimal levels of emissions for a nonsignatory j and a signatory i in Period 1 are

$$\begin{aligned} e_{j,1} &= \left[\gamma \left(1 + \beta\delta \right) \right]^{-1/(1-b)} \quad (3.27) \\ e_{i,1} &= \left[\gamma n_1 \left(1 + \beta\delta \right) \right]^{-1/(1-b)} \quad (3.28) \end{aligned}$$

The level of emissions of a signatory i is less than that of a nonsignatory j. Signatories would cut more emissions when more countries are in the coalition. The result also shows that the higher discount factor (β) and the higher emis-

sion decay rate (δ) would also lead to a lower level of emissions. It means if the young generation is more valuable to the old generation, the decision makers in Period 1 would do more abatement for the sake of the young generation in Period 2.

Taking the expected number of signatories n_1 into (3.6), the expected level of emission for a signatory i in Period 2 is higher than the level for a signatory i in Period 1. The level of emissions for nonsignatory j in Period 2 is also higher than that in Period 1. Compared to the result in the myopic scenario which the levels of emissions are the same for two periods, the old generation would do more abatement for the young generation in the sustainable development scenario.

Case 2. The sustainability criterion for signatories is binding ($\lambda_j = 0$, but $\lambda_i > 0$)

When the sustainability criterion for nonsignatories is not binding, $\lambda_j = 0$. From (3.20), the level of emissions for a nonsignatory j is

$$e_{j,1} = \left[\gamma\left(1+\beta\delta\right)\right]^{-1/(1-b)} \tag{3.29}$$

On the other hand, when the criterion is binding for signatories, we assume $\lambda_i > 0$. The level of emissions of a signatory i can be derived from (3.26)

$$\begin{aligned} & \frac{1}{b}e_{i,1}^{b} - \gamma\left(1-\delta\right)\left[\delta E_0 + n_1 e_{i,1} + \left(N-n_1\right)e_{j,1}\right] \qquad (3.30) \\ = \; & \frac{1}{b}\left(n_1\gamma\right)^{\frac{-b}{1-b}} - \gamma\left[n_1\left(n_1\gamma\right)^{\frac{-1}{1-b}} + \left(N-n_1\right)\left(\gamma\right)^{\frac{-1}{1-b}}\right] \end{aligned}$$

Suppose countries have a high discount rate ($\beta \approx 1$) and the remaining

emissions is high ($\delta \approx 1$), an old nonsignatory emits $[2\gamma]^{-1/(1-b)}$ which is less than the result in the MYO scenario. When the sustainability criterion for signatories is binding, from (3.30) we learn that the level of emission for an old signatory $\{(n_1\gamma)^{\frac{-b}{1-b}} - b\gamma\left[n_1\left(n_1\gamma\right)^{\frac{-1}{1-b}} + \left(N-n_1\right)\left(\gamma\right)^{\frac{-1}{1-b}}\right]\}^{1/b}$ which is the benefit elasticity times the expected welfare of a young signatory to the power of the inverse benefit elasticity of emissions b. On the other hand, when countries have a low discount rate ($\beta \approx 0$) or the remaining emissions is low ($\delta \approx 0$), an old nonsignatory emits $[\gamma]^{-1/(1-b)}$ which is at the same level to the result in the MYO scenario. Because the remaining emissions leads to an extra cost to the young generation, an old signatory emits less than the expected welfare of a young signatory.

Case 3. The sustainability criterion for nonsignatories is binding ($\lambda_j > 0$, but $\lambda_i = 0$)

When the sustainability criterion for nonsignatories is binding, $\lambda_j > 0$. The level of emissions of a nonsignatory j can be derived from (3.21)

$$\begin{aligned} & \frac{1}{b}e_{j,1}^{b} - \gamma\left(1-\delta\right)\left[\delta E_0 + n_1 e_{i,1} + \left(N-n_1\right)e_{j,1}\right] \qquad (3.31) \\ = \; & \frac{1}{b}\left(\gamma\right)^{\frac{-b}{1-b}} - \gamma\left[n_1\left(n_1\gamma\right)^{\frac{-1}{1-b}} + \left(N-n_1\right)\left(\gamma\right)^{\frac{-1}{1-b}}\right] \end{aligned}$$

On the other hand, if the criterion for signatories is not binding, $\lambda_i = 0$. From (3.25), the level of emissions of a signatory i is therefore

$$e_{i,1} = \left[\gamma n_1\left(1+\beta\delta\right)\right]^{-1/(1-b)} \qquad (3.32)$$

Suppose countries have a high discount rate ($\beta \approx 1$) and the remaining emissions is high ($\delta \approx 1$), an old signatory emits $[2\gamma n_1]^{-1/(1-b)}$ which is less than the result in the MYO scenario. When the sustainability criterion for nonsignatories is binding, from (3.31) we learn that the level of emission for an old nonsignatory $\{(\gamma)^{\frac{-b}{1-b}} - b\gamma\left[n_1\left(n_1\gamma\right)^{\frac{-1}{1-b}} + (N-n_1)\left(\gamma\right)^{\frac{-1}{1-b}}\right]\}^{1/b}$ which is the benefit elasticity times the expected welfare of a young nonsignatory to the power of the inverse benefit elasticity of emissions b. When countries have a low discount rate ($\beta \approx 0$) or the remaining emissions will be very low ($\delta \approx 0$), an old signatory emits $[\gamma n_1]^{-1/(1-b)}$ which is at the same level to the result in the MYO scenario. Because the criterion is active and the remaining emissions leads to an extra cost to the young generation, an old nonsignatory emits less than the expected welfare of a young nonsignatory.

Case 4. The sustainability criteria for all countries are binding ($\lambda_j > 0$, $\lambda_i > 0$)

In this case, $\lambda_j > 0$ and $\lambda_i > 0$. The levels of emissions of a nonsignatory j and a signatory i can be derived from (3.21) and (3.26) as

$$\begin{aligned} & \frac{1}{b}e_{j,1}^{b} - \gamma\left(1-\delta\right)\left[\delta E_0 + n_1 e_{i,1} + (N-n_1)\, e_{j,1}\right] \\ = \; & \frac{1}{b}\left(\gamma\right)^{\frac{-b}{1-b}} - \gamma\left[n_1\left(n_1\gamma\right)^{\frac{-1}{1-b}} + (N-n_1)\left(\gamma\right)^{\frac{-1}{1-b}}\right] \end{aligned}$$

$$\begin{aligned} & \frac{1}{b}e_{i,1}^{b} - \gamma\left(1-\delta\right)\left[\delta E_0 + n_1 e_{i,1} + (N-n_1)\, e_{j,1}\right] \\ = \; & \frac{1}{b}\left(n_1\gamma\right)^{\frac{-b}{1-b}} - \gamma\left[n_1\left(n_1\gamma\right)^{\frac{-1}{1-b}} + (N-n_1)\left(\gamma\right)^{\frac{-1}{1-b}}\right] \end{aligned}$$

The discount factor (β) affects neither a signatory nor a nonsignatory.

The remaining level of emissions (δ) is an important factor when the decision makers decide the level of emissions. When the remaining emissions will be very small ($\delta \approx 0$), the pollution will be absorbed by the nature. The old generation would emit at the level as the benefit elasticity of emissions b times the expected welfare of the young generation to the power of the inverse b. But if the nature cannot absorb the pollution and the remaining emissions is at a very high level ($\delta \approx 1$), the old generation has to emit less if they consider the cost of the accumulated emissions to the young generation.

The optimal levels of emissions for a signatory and a nonsignatory are not obvious. A numerical example in the following section can illuminate the results in these cases.

First-stage membership game

To find a stable coalition in the first period, we rewrite the internal constraint and external constraint for the old generation as

$$\pi_{j,1}\left(n_1^* - 1\right) + \beta\pi_{j,2}^f\left(n_1^* - 1\right) \leq \pi_{i,1}\left(n_1^*\right) + \beta\pi_{i,2}^f\left(n_1^*\right) \tag{3.33}$$

$$\pi_{i,1}\left(n_1^* + 1\right) + \beta\pi_{i,2}^f\left(n_1^* + 1\right) \leq \pi_{j,1}\left(n_1^*\right) + \beta\pi_{j,2}^f\left(n_1^*\right) \tag{3.34}$$

The constraints with a cross-generational objective function imply that the decision makers take the expected welfare of the young generation into account. The constraint (3.33) shows that when the welfare of being a nonsignatory is not higher than that of being a signatory, the coalition is stable internally. On the other hand, the constraint (3.34) shows that the coalition is stable externally, when there is no signatory have the incentive to leave.

Consider the case where $n_1 = N$ where all countries join the IEA, the indi-

vidual levels of emissions are $[\gamma N(1+\beta\delta)]^{-1/(1-b)}$ in Period 1 and $(\gamma N)^{-1/(1-b)}$ in Period 2. The expected level of emission in Period 2 is higher than that in Period 1. This implies that the old generation has lower benefit to the young generation, however, the cost for the old generation is also smaller. It is unclear to say whether this is a sustainable system. Hence, the following simulation provides a numerical example to illuminate the result.

3.3 Simulation analysis

Given $N = 10$ countries[9], we assume the gap between generations is five decades because the international treaties are usually valid for a long term. The decay rate of total emissions (δ) is set as $(100 - 0.866)\%$ per year from the natural annual removal rate of CO2 stock given by Nordhaus (1994). The parameters of benefit (b) is set from 0.01 to 0.1 and the marginal cost of total emissions (γ)[10] is set from 0.01 to 0.9.

Table 3.2 shows the individual level of emissions and welfare in the myopic (MYO) scenario. As mentioned previously, a signatory produces less pollution than a nonsignatory does. Hence, the payoff of a signatory is less than that of a nonsignatory in both periods. The individual optimal emissions levels of signatories and nonsignatories in two different periods are positively affected

[9]We acknowledge that $N = 10$ might not a large number, compared to the numerical examples in Barrett (1994) and Rubio and Ulph (2007). It is more difficult to find a robust result in our exponential benefit function with a case of large number of countries. Hence, this assumption is adequate to represent an international negotiation while a robust result could be found.

[10]Here we assume the marginal cost is at the range of 0 and 1. As we mentioned in footnote 5, when the marginal cost γ is less than 1, the higher technology will increase the emission level. It implies that when the technology efficiency improvement is faster than the increasing cost, the overall emission will increase.

γ	b = 0.01	b = 0.02	b = 0.05	b = 0.1
0.02	$\begin{pmatrix} 52.02 & 94.68 \\ 25.83 & 93.95 \\ 52.02 & 93.78 \\ 10.24 & 92.10 \end{pmatrix}$	$\begin{pmatrix} 54.16 & 44.42 \\ 26.70 & 43.66 \\ 54.16 & 43.51 \\ 10.48 & 41.76 \end{pmatrix}$	$\begin{pmatrix} 61.43 & 13.56 \\ 29.62 & 12.68 \\ 61.43 & 12.59 \\ 11.29 & 10.60 \end{pmatrix}$	$\begin{pmatrix} 77.22 & 1.66 \\ 35.75 & 0.51 \\ 77.22 & 0.60 \\ 12.92 & 0.00 \end{pmatrix}$
0.1	$\begin{pmatrix} 10.24 & 93.15 \\ 5.08 & 92.43 \\ 10.24 & 92.26 \\ 2.01 & 90.61 \end{pmatrix}$	$\begin{pmatrix} 10.48 & 42.99 \\ 5.17 & 42.25 \\ 10.48 & 42.10 \\ 2.03 & 40.41 \end{pmatrix}$	$\begin{pmatrix} 11.29 & 12.46 \\ 5.44 & 11.65 \\ 11.29 & 11.57 \\ 2.07 & 9.74 \end{pmatrix}$	$\begin{pmatrix} 12.92 & 1.39 \\ 5.98 & 0.43 \\ 12.92 & 0.50 \\ 2.16 & 0.00 \end{pmatrix}$
0.5	$\begin{pmatrix} 2.01 & 91.65 \\ 1 & 90.94 \\ 2.01 & 90.78 \\ 0.40 & 89.15 \end{pmatrix}$	$\begin{pmatrix} 2.03 & 41.60 \\ 1 & 40.89 \\ 2.03 & 40.74 \\ 0.39 & 39.10 \end{pmatrix}$	$\begin{pmatrix} 2.07 & 11.45 \\ 1 & 10.70 \\ 2.07 & 10.63 \\ 0.38 & 8.95 \end{pmatrix}$	$\begin{pmatrix} 2.16 & 1.16 \\ 1 & 0.36 \\ 2.16 & 0.42 \\ 0.36 & 0.00 \end{pmatrix}$
0.9	$\begin{pmatrix} 1 & 91.01 \\ 0.50 & 90.31 \\ 1 & 90.14 \\ 0.20 & 88.53 \end{pmatrix}$	$\begin{pmatrix} 1 & 41.01 \\ 0.49 & 40.31 \\ 1 & 40.17 \\ 0.19 & 38.55 \end{pmatrix}$	$\begin{pmatrix} 1 & 11.04 \\ 0.48 & 10.32 \\ 1 & 10.25 \\ 0.18 & 8.63 \end{pmatrix}$	$\begin{pmatrix} 1 & 1.07 \\ 0.46 & 0.33 \\ 1 & 0.39 \\ 0.17 & 0.00 \end{pmatrix}$

Given $N = 10$ and $\delta = (1 - 0.00866)^{50}$. From left top to down in each cell are the emissions of a nonsignatory and a signatory in period 1 and a nonsignatory and a signatory in period 2 respectively in the MYO scenario. From right top to down are their individual payoffs.

Table 3.2: Individual level of emissions and welfare of a nonsignatory and a signatory in two periods in the myopic (MYO) scenario

by the technology level (b) and negatively affected by the marginal cost of total emissions (γ).

The membership decision is determined *ex ante* the emissions game. A consistent result in the MYO scenario is that there is always a 2-member coalition in Period 1, and a larger 5-member coalition in Period 2. The individual level of emissions and welfare are affected by the size of IEA. The nonsignatories generate the same level of emissions in two periods, while the signatories emit less in Period 2. When the payoffs between generations are compared, the old generation has a higher payoff than the young generation. In other words, the system in the MYO scenario is always unsustainable.

The results are summarised as follows.

Summary 6 *In the myopic (MYO) scenario,*
(1) Nonsignatories emit the same quantity in both periods. The old signatories emit more than the young signatories. There is no fairness between generations, the old generation always has higher welfare than the young generation.
(2) The level of emissions is higher when the technology is more developed. The welfare is therefore lower with the more advanced technology. On the other hand, the higher the marginal cost of the total emissions is, the lower the level of emissions and welfare will be.
(3) Few countries have the incentives to participate in an IEA in Period 1*, compared to the outcome in Period* 2*.*

Table 3.3 reports the individual level of emissions and welfare in the sustainable development (SD) scenario. Here, the discount rate for the next generation (β) is set as 0.5. The level of emissions in Period 1 is less than that

γ	b = 0.01	b = 0.02	b = 0.05
0.02	— — / 3.83 93.85 / — — / 5.08 95.77	— — / 3.88 43.59 / — — / 5.17 45.60	— — / 4.05 12.69 / — — / 5.44 15.01
0.05	— — / 1.52 92.99 / — — / 0.05 94.89	— — / 1.52 42.78 / — — / 2.03 44.75	— — / 1.54 12.09 / — — / 2.07 14.30
0.5	1.52 87.84 / 0.50 86.73 / 0.50 92.14 / 0.25 91.44	1.52 37.78 / 0.50 36.66 / 0.49 42.12 / 0.24 41.43	1.54 7.59 / 0.49 0.44 / 0.48 12.08 / 0.23 11.39
0.6	1.26 87.68 / 0.42 86.57 / 0.60 90.41 / 0.27 89.64	1.27 37.61 / 0.41 36.53 / 0.59 40.41 / 0.27 39.64	1.27 7.52 / 0.40 6.38 / 0.58 10.40 / 0.26 9.62
0.9	0.13 88.72 / 0.43 89.88 / 0.90 89.66 / 0.18 88.08 *	0.13 38.74 / 0.43 39.89 / 0.90 39.70 / 0.18 38.11 *	0.12 8.79 / 0.42 9.95 / 0.90 9.8 / 0.17 8.21 *

Given $N = 10$, $\delta = (1 - 0.00866)^{50}$ and β is 0.5. From left top to down in each cell are the emissions of a nonsignatory and a signatory in Period 1 and a nonsignatory and a signatory in Period 2 respectively in the SD scenario. From right top to down are their individual payoffs in Periods 1 and 2.

The cells with star * refer to the sustainability criterion is binding.

Table 3.3: Individual emission levels and the welfare of a nonsignatory and a signatory in two periods in the sustainable development (SD) scenario

in Period 2 in general. When the technology is more advanced (higher b), the emissions level increases but the welfare shrinks. On the other hand, when the marginal cost of the total emissions (γ) increases, countries are more aware of the damage and reduce the levels of emissions. The marginal cost has positive effect on the emissions level but negative effect on the welfare.

The cells with star refer to the binding sustainability criterion that the expected welfare in Period 2 are worse than that in Period 1. The system could be sustainable in most cases, but not always. We have to emphasise that the sustainability criterion is for the old generation in Period 1. When the criterion is binding, the expected welfare in Period 2 is equal to the welfare in Period 1. However, due to the coalition formation might be changed by the young generation in Period 2, the actual welfare in Period 2 is not necessary to be the expected welfare. The numerical example shows that the criteria are not binding when the marginal cost of total emissions is high. In the SD scenario, the system is usually sustain that the welfare of the young generation are higher than the welfare of the old generation. However, when the marginal cost is high, the system could be unsustain that young generation might yield a lower level of welfare.

Compared to the result in the MYO scenario in Table 3.2, the level of emissions of SD scenario is far less than that of MYO scenario. In addition, the welfare of signatories and nonsignatories in Period 2 in the SD scenario are usually higher than those in the MYO scenario. In other words, the SD scenario is better to maintain a sustainable system than the MYO scenario.

Table 3.4 reports the coalition formation of IEAs in the SD scenario. When the marginal cost of the stock of emissions (γ) is low, the grand coalition

γ	b		
	0.01	0.02	0.05
0.02	10	10	10
	10	10	10
0.05	10	10	10
	10	10	10
0.5	3	3	3
	8	8	8
0.6	3	3	3
	6	6	6
0.9	6	6	6
	6	6	6

The discount rate (β) is 0.5. From top to down in each cell report the number of signatories in the periods 1 and 2.

Table 3.4: Number of signatories out of 10 for the parameter of the level of technology and the marginal cost of the total emissions in the SD scenario

could be formed. Countries have a higher incentive to form an IEA when the marginal cost is low. When the marginal cost increases, the coalition formation in Period 2 decreases. However, the marginal cost has ambiguous impact on the formation in Period 1. Compared to the result in MYO scenario where there are always a 2-member coalition in Period 1 and 5-member in Period 2, the formation in the SD scenario is larger than that in the MYO scenario. On the other hand, the level of technology (b) has no impact on the coalition formation in the SD scenario, while there is also no impact in the MYO scenario.

Table 3.5 shows the sizes of stable IEAs in the SD scenario in relation to the levels of discount rate (β) and the marginal cost of total emissions (γ) when the technology level b is set at 0.05. A grand coalition exist when the marginal cost of total emissions is very low. When the marginal cost increases, grand coalition does not exist. However, the marginal cost does not show a clear correlation with the coalition formation in two periods. It seems that the

γ	β				
	0.01	0.25	0.5	0.75	1
0.02	10	10	10	10	10
	10	10	10	10	10
0.4	2	3	3	4	4
	10	10	10	10	10
0.5	2	3	3	3	3
	10	9	8	8	7
0.9	6	6	6	6	6
	6	6	6	6	6
From top to down in each cell report the number of signatories in the periods 1 and 2					

Table 3.5: Number of signatories out of 10 for the parameter of the perceptions of sustainability and the marginal cost of the total emissions in the SD scenario (b=0.05)

formation in Period 2 decreases when the marginal cost increases, while that in Period 1 may firstly shrink then expanded. When the discount rate (β) is very small, it implies that the old generation's preference weighting attached by one generation to the welfare of the next, the formation in Period 1 could be very small but a grand coalition is still possible in Period 2. It is interesting that the discount rate has small but ambiguous effect on the coalition formation.

We have to note that a robust outcome is not found when the level of discount rate s more than 0.05, however, the impact of the discount rate is not as huge as the marginal cost of total emission. The coalition formation usually increases when the marginal cost grows.

The results are summarised as follows.

Summary 7 *In the sustainable development (SD) scenario,*

(1) When the marginal cost of the total emissions increases, countries are more aware of the damage and will reduce the levels of emissions. The individual welfare therefore increases. Besides, countries have higher incentives to par-

ticipate in an IEA when the cost is low. A grand coalition is possible in the SD scenario.

(2) When the level of technology development is more advanced, the levels of individual emissions increases and the payoffs are smaller. The coalition size is no change to a different developed technology.

(3) When countries have higher discount rate to the welfare of the next generation, the coalition formation may increase in Period 1 *but decrease in Period* 2*. However, the impact on the formation is small.*

(4) The sustainability criterion is usually binding when the marginal cost of the total emissions is low. The old generation would emit less when the criterion is binding. If the technology development is more developed, each country receives higher welfare compared to the outcomes in the myopic scenario.

3.4 Conclusions

This chapter examines the effect of preference weighting attached by one generation to the welfare of the next and the sustainability criterion on the formation of IEAs. To do so, we have built a model with a two-stage game in two periods to examine the impact of the discount rate and the sustainability criterion on the formation of international environmental agreements. We firstly consider a myopic (MYO) scenario in which the old generation is myopic and does not care about the young generation. It implies that there is no fairness and altruism between generations. The old generation only concerns about their current payoff in Period 1. The result shows that, only a small size (2 members) coalition could possibly be formed in Period 1 and a larger (5 members) coalition in Period 2. The simulation results show that the framework of an IEA remains

unchanged given the level of marginal cost of the total emissions and the level of technology development. The level of emissions decreases when the marginal cost increases. On the other hand, a more advanced technology development level could encourage countries to emit more and have lower welfare. The system in the MYO scenario is demonstrated to be unsustainable.

This study then builds a model in the sustainable development (SD) scenario which is characterised of two intergenerational behaviours. Firstly, the countries have intergenerational altruism; they care about not only their welfare in Period 1 but also that of the young generation in Period 2. Secondly, the countries care about the intergenerational fairness whereby the old generation should not make the young generation worse off. The simulation results show that a grand coalition is possibly formed when the marginal cost of the total emissions is very low. But the impact of the discount rate is small and performs differently in two periods. On the other hand, the technology development level has no impact on the formation. The sustainability criterion is binding usually when the marginal cost is high, the old generation has to reduce the emissions level in order to ensure the sustainability criterion is binding. The young generation usually has better welfare than the old generation has. However, it must be noted that the criterion does not guarantee a sustainable system. In a few cases, the system are still unsustainable because the young generation could make decisions different to what the old generation expected.

This study confirms the importance of the awareness of sustainability on IEAs formation. When the intergenerational fairness and altruism are taken into account, a formation will be expanded. Besides, the marginal cost of the total emissions is an important factor for the formation of IEAs. The advanced

level of technology development may lead a more efficient production per unit of emission, but it also encourages countries to emit more and have a lower level of welfare.

Conclusions

Typically, the studies on international environmental agreements (IEAs) are based on the assumption of egoistic agents. They attempt to find a Nash equilibrium where all agents seek to maximise their own monetary payoffs. However, the existing laboratory evidences on IEAs often yield 'noisy' results in the sense that individual choices are difficult to map because their preferences are complicated and not always egoistic. This thesis has investigated individual behaviours and incentives related to the economics of IEAs through experiments and numerical examinations. The thesis as a whole contributes to the economics of international environmental agreements by providing both theoretical and experimental perspectives. More specifically, the thesis takes into account heterogeneity of players in IEAs and provides a deeper understanding of social preferences (of fairness, altruism and sustainability) in not only a static model in chapters 1 and 2 but also a cross-generational model in chapter 3.

Both Chapter 1 and Chapter 2 have investigated factors that shape decision-making in a static model which determines the coalition formation. The adoption of an assumption of heterogeneous players and experimental methods

advances the academic debate because they provide more realistic current explanation. Chapter 1 has examined the impacts of inequality-averse preferences on individual decisions, and even on IEA coalition formation. Chapter 2 has examined the impacts of altruistic preferences. Both chapters 1 and 2 have provided experimental evidences on social preferences and how they shape decision-making processes in a static public good game. Although there exist some internal and external constraints which influence the stability of coalition formation, these two chapters aim to identify a particular unique equilibrium condition so as to access individual preferences. As mentioned in the Introduction, the assumption of a self-interested preference is not robust enough to explain low free-ride effect in a public good game. If we can more specifically identify individual social preferences, we can better understand how a stable coalition can be formed.

In order to scrutinise the theories in each chapter, experimental methods and numerical simulation have been adopted. Chapters 1 and 2 have employed novel experimental designs to determine individual social preferences. Experimental methods have a great ability to capture individual heterogeneous behaviours which is one of the main assumptions in our models. The subjects in the experiments take a series of decisions which indicated their individual preferences on inequality-aversion and altruism. We test the existing theories that consider social preferences as important determinants for motivating participation in an IEA (Kolstad, 2014; Hahn and Ritz, 2014) by introducing eight particular treatments which are able to capture the subjects' individuality. In order to capture individual behaviours in IEAs, our model has been designed with a unique equilibrium condition. Each subject has a weakly dominant strategy of whether or not to join a coalition. In contrast to

what the literature suggests, the results from this particular experiment design do not support the necessity that higher marginal benefit would enlarge the coalition size and the total contribution. Instead, it illustrates the formation is conditional on the combination of all agents' marginal benefits.

Chapter 1 that analyses the impact of inequality-averse preferences on the formation of IEAs has suggested that coalition formation could be either an unstable coalition, a stable coalition as suggested by the Nash prediction, a stable coalition which is larger than the Nash prediction, depending on the degrees of inequality-averse preferences. On the one hand, when one signatory is strongly inequality-averse, the internal constraint may be violated and signatories may leave the coalition. On the other hand, when one nonsignatory has strong attitude to fairness, the external constraint will always be hold and nonsignatories will prefer to have the free-riding benefit and be absent from the coalition.

Although the experimental evidences in chapter 1 confirm the impact of inequality-averse preferences on coalition formation, the experimental outcomes do not support the theoretical prediction. On the contrary, it has shown that subjects with lower degree of inequality-aversion are more likely to act strategicly by breaking the internal constraint. By doing so, they could force free-riders to participate. When their role is switched to the opposite side, they play strategicly by compromising their free-riding payoffs. Some other variables in the questionnaire, such as the political preference and religion preference, have had a significant impact on the subjects' decisions. For example, pro-right-wingers behave as those with low degree of inequality-aversion and make some strategic decisions.

Based on the findings, Chapter 1 has shed light on the policy-making of IEAs: In order to stabilise a coalition internally, the international bodies had better emphasise the importance of fairness to signatories. Non-signatories may feel threatened by potential damages if the international bodies fail to achieve the targets of an IEA. In this regard, countries have higher willingness to participate in an IEA. Although this study confirms the existence of the inequality-averse preferences, it does not distinguish disadvantage loss from advantage loss. It is intuitive to assume that advantage loss is smaller than disadvantage loss, but it needs to be proved by experiments. Therefore, an experiment that can indicate individual inequality-averse preferences can be further developed in the future. However, there is a challenge of how to accurately capturing individual preferences in this experiment.

Chapter 2 examines the impact of altruistic preference on the IEA formation. The theoretical hypothesis states that agents who have high degree of altruistic preferences are more likely to cooperate. Depending on individual attitudes to altruism, coalition formation could be either a stable coalition as the Nash prediction suggests, or a stable coalition larger than that. All signatories with a high degree of altruistic preference have no incentive to violate the internal constraints by leaving a coalition. On the other hand, a subject who has a high degree of altruistic preference may violate the external constraint. The outcome of this is stronger cooperation in the coalition formation.

However, the experimental evidences show that altruistic preference is a significant negative factor to the incentives to participate in a coalition. This is contrast to our theoretical prediction. The experimental results show that the subjects' behaviours are strategic in an interactive game . This is particularly

so where subjects' weakly dominant are not to join a coalition.

Future work could explore intergenerational altruism by using alternative approaches. For instances, Karp (2013) proposes an overlapping generation (OLG) framework with intergenerational altruism integrated into a differential game between nations. By comparing analytic results for a linear model and numerical results for a convex model, he argues that the importance of altruism depends on model specifics and the equilibrium type. Future studies could examine his arguments using experimental methods.

Another future work for the studies of social preferences would be to explore the link between individual preferences and behaviours. Strategic behaviours have been observed in our experiments, but the reasons and the processes have not been understood. Hahn and Ritz (2014) also find that it may be difficult to infer countries' true preferences for altruism from their observed behaviour. The challenge to future studies on climate negotiation is to find out the link between players' preferences and behaviours. Reciprocation could be an interesting field to be explored (Hwang and Bowles, 2012; Hadjiyiannis *et al.*, 2012a, 2012b). Positive reciprocity refers to the situation where countries receive mutual benefits and get reward for fair behaviour, whereas negative reciprocity refers to that when countries retaliate against each other and behave unfairly. When subjects have high degree of social preferences (either altruistic or inequality-averse), they might expect that other subjects have similar moral standard and act strategicly. Future studies can examine this hypothesis by using experimental evidences.

Another possible perspective is that, individual preferences could vary over time. Matros (2012) analyses an evolutionary version of the public good game in which boundedly rational agents can use imitation and best-reply decision

rules. The dynamics of preferences would be another task for future studies.

Chapter 3 has investigated the discount factor attached by one generation to the welfare of the next and how these different understandings of sustainability shape from different generation's individual decisions. At international conventions on climate change, sustainability is one of the most important reasons often quoted to form an IEA. However, it has not been discussed extensively in the literature. Chapter 3 has bridged this gap by investigating the role of the discount rate attached by one generation to the next generation. To this end, a numerical simulation where some parameters selected from existing scientific evidences has been built. We consider a two-generation model with a two-stage game in two periods to examine the impact of the discount rate and the sustainability criterion on the formation of international environmental agreements. Decision makers live in one period: the old generation live in Period 1 and the young generation live in Period 2. Each generation faces a two-stage decision. In the first stage membership game, each country decides whether or not to join an IEA. The coalition formation is determined in this stage. In the second stage, in terms of their membership status, they decide their individual levels of emissions.

In order to examines the effect of discount rate and the sustainability criterion on the formation of IEAs, two scenarios are built: myopic (MYO) scenario and sustainable development (SD) scenario. In the MYO scenario, countries are myopic in the sense that the old generation cares about its welfare only. In the SD scenario, the old generation cares not only its welfare but also the welfare of the young generation. Besides, the old generation should not make the young generation worse off. These two main features in the SD scenario

to represent the concepts of sustainability: first, we assume that the present generation has cross-generational altruism on future generations. Second, we assume that the present generation concerns cross-generational fairness that the welfare of the future generation is no worse than that of the current generation.

In the MYO scenario, there is no fairness and altruism between generations. The old generation only cares about their current payoff in Period 1. The numerical result shows that, only a small size (2 members) coalition could possibly be formed in Period 1 and a larger (5 members) coalition in Period 2. The simulation results show that the framework of an IEA remains unchanged given the level of marginal cost of the total emissions and the level of technology development. The level of emissions decreases when the marginal cost increases. On the other hand, a more advanced technology development level could encourage countries to emit more and have lower welfare. The system in the MYO scenario is demonstrated to be unsustainable.

Then, we consider the preference weighting for the next generation and the sustainability criterion in the sustainable development (SD) scenario. We assume that the old generation cares not only its welfare but also the welfare of the young generation. Besides, the sustainability criterion requires the old generation should not make the young generation worse off. The simulation results show that a grand coalition is possible in two periods when the marginal cost of the total emissions is very low. But the impact of the discount rate is small and different on two periods. On the other hand, the technology development level has no impact on the formation. The sustainability criterion is binding usually when the marginal cost is high, the old generation has to reduce the emissions level in order to ensure the sustainability criterion is

binding. The young generation usually has better welfare than the old generation has. However, it must be noted that the criterion does not guarantee a sustainable system. In a few cases, the system are still unsustainable because the young generation could make decisions different to what the old generation expected.

This study confirms the importance of the awareness of sustainability on IEAs formation. When the intergenerational fairness and altruism are taken into account, a formation will be expanded. Besides, the marginal cost of the total emissions is an important factor for the formation of IEAs. The advanced level of technology development may lead a more efficient production per unit of emission, but it also encourage countries to emit more and have a lower level of welfare.

Future work could replace the existing assumption of homogeneity by adding heterogeneous preference to the model. So far, Chapter 3 only examines a basic model that captures the notion of sustainability. As mentioned in Chapter 1, the assumption of homogeneous preferences limits our understanding of the reality. In addition, the impact of different discounts rate on policy mechanism can be explored further. For example, Carraro *et al.* (2009) have considered minimum participation constraint which is a frequent mechanism in environmental treaties. This mechanism is designed to reduce the free-riding effect in a public good game. The minimum participation rule can be considered in the future study. In addition, the two-generation model may not a very satisfactory model to study the impact of the non-declining welfare constraint. Other designs, e.g. an infinite-horizon model or over-lapping generations model, may be better suited.

To conclude, this thesis has illustrated individual incentives of participating in IEAs as well as the coalition formation through both experimental and theoretical findings. In static models, this thesis claim that the formation of IEAs is affected by individual social preferences. However, the experimental evidence suggests that individual decisions are far more complex than our theoretical predictions with a single type of preference (e.g. fairness or altruism). Subjects in the experiment behaved strategicly in the individual and interactive games. Furthermore, their subjective attitudes to politics and religion also play an important role in the willingness to participate. Whereas the finding contrasts with the intuition that left-wingers and religionists are traditional supporters in practical IEAs.

In a two-period model, this thesis has examined the impact of discount rate, which are considered as cross-generational social preferences, on the coalition formation. This study confirms the importance of the awareness of sustainability to international environmental conventions. When the intergenerational fairness and altruism are taken into account, a coalition formation will be expanded. The numerical example indicates that the marginal cost of the total emissions is an important factor for the formation of IEAs. In contrast, the advanced level of technology development may lead a more efficient production per unit of emissions, but it also encourages countries to emit more in total and have a lower level of welfare. Only when the marginal cost to total emissions is low and the current generation concerns the future generation, a sustainable system could be succeed.

Appendices

Appendix 1.1

Proof. To proof the theorem, we establish an algorithm to find a stable coalition. Player n^* has the incentive to maintain n^*-member coalition if the payoff $\pi^s_{n^*}(n^*) = (-1) + \sum_i^{n^*} \gamma_i$ is positive. If player n^* leaves, the coalition collapsed. Hence, player n^* gets $\pi^{ns}_n(n-1) = 1$ when all player pollute. When the internal constraint makes player n^* to be stable in the coalition, all signatories have the same incentive to make it stable internally.

Meanwhile, the external constraint asks player N to stay away from the n^*-member coalition. When player N is a nonsignatory, its payoff is $\pi^{ns}_N(n^*) = (\gamma_N \cdot n^*)$. If player N changes its mind and joins the coalition as the (n^*+1)-th member, the payoff becomes $\pi^s_N(n^*+1) = \left[(-1) + \sum_i^{n^*} \gamma_i\right] + \gamma_N$. When the external constraint deters player N to join the coalition, all nonsignatories are deterred and the coalition becomes stable externally. Hence, the theorem is established.

By the internal and external constraints, the minimum number to form a profitable coalition is found. However, this coalition is not the only equilibrium. A coalition with more members could be another equilibrium if and only if both constraints are held. A unique equilibrium exists when any member is irreplaceable by a larger coalition. It means that, if all nonsignatories would like to replace the player n^* with the smallest marginal benefit of abatement (γ_{n^*}) in the coalition, the coalition would collapse. In other words, a $(N-1)$-

member without player n^* is unprofitable. We can write it in an inequality

$$1 > \sum_{i=1}^{n^*-1} \gamma_i + \sum_{j=n^*+1}^{N} \gamma_j$$

To add the marginal benefit of abatement of player n^* in both sides, the unique equilibrium condition is rewritten as

$$1 + \gamma_{n^*} > \sum_{i=1}^{N} \gamma_i$$

■

Appendix 1.2

Proof. The utility of a signatory i with a n-member coalition can be extended to the function with the degree of inequality-aversion as

$$u_i^s(n) = \pi_i^s(n) - \frac{\alpha_i}{N-1}\sum_{m\neq i}\max\left[\pi_m - \pi_i^s(n), 0\right] \tag{3.35}$$

Because of the external constraint, any nonsignatory has higher utility than what a signatory has. Signatory i has the disadvantage term but no advantage term.

On the other hand, the welfare function of a nonsignatory j with n-member coalition is

$$u_j^{ns}(n) = \pi_j^{ns}(n) - \frac{\alpha_j}{N-1}\sum_{j\neq m}\max\left[\pi_m - \pi_j^{ns}(n), 0\right] - \frac{\beta_j}{N-1}\sum_{m\neq j}\max\left[\pi_j^{ns}(n) - \pi_m, 0\right] \tag{3.36}$$

Nonsignatories could have both the advantage and disadvantage terms. They are advantaged since their individual payoffs are definitely higher than that of a signatory. The one with the highest marginal benefit of the total abatement yields the highest payoff among others. Any other nonsignatory would be disadvantaged to this country.

The stability of the coalition formation depends on the internal and the

external constraints. The internal one can be displayed as

$$
\begin{aligned}
& u_i^s\left(n^*\right) > u_i^{ns}\left(n^*-1\right) \\
\Longrightarrow \left(-1+\sum_{i=1}^{n^*}\gamma_i\right) - \frac{\alpha_i}{N-1}\sum_{j=n^*+1}^{N}\left[n^*\gamma_j - \left(-1+\sum_{i=1}^{n^*}\gamma_i\right)\right] & > 0
\end{aligned}
$$

The left-hand-side is the utility when i joins the coalition, and the right-hand-side is the utility when i does not join.

If i is not strong inequality averse, the player would follow the internal constraint and decide to participate the coalition. If i is strong inequality averse, both the individual inequality-averse factor α_i and the disadvantage loss are high enough, the player would violate the internal constraint and the consequence is a collapse coalition.

On the other hand, the external constraint can be extended as

$$
\begin{aligned}
& u_j^{ns}\left(n^*\right) > u_i^s\left(n^*+1\right) \\
\Longrightarrow \quad & n^*\gamma_k - \frac{\alpha_k}{N-1}\sum_{k\neq j}\max\left[\pi_j - \pi_k^{ns}\left(n^*\right), 0\right] - \frac{\beta_k}{N-1}\sum_{k\neq j}\max\left[\pi_k^{ns}\left(n^*\right) - \pi_j, 0\right] \\
> \quad & \left(-1+\sum_{i=1}^{n^*}\gamma_i+\gamma_k\right) - \frac{\alpha_k}{N-1}\sum_{k\neq j}\max\left[\left(n^*+1\right)\gamma_j - \left(-1+\sum_{i=1}^{n^*}\gamma_i+\gamma_k\right)\right]
\end{aligned}
$$

where k is a player belongs to $[n^*+1, N]$. The left-hand-side is k's utility when k is a nonsignatory and have the disadvantage loss from higher marginal benefit nonsignatories as well as the advantage loss from all signatories and lower marginal benefit nonsignatories. The right-hand-side is k's utility when k is a signatory which only has the disadvantage loss.

When k does not have enough advantage averse, the player would follow the external constraint and not to participate in the coalition. When k has

strong inequality aversion, both the individual inequality-averse factor α_k and β_k, and the disadvantage and advantage loss are high, the player would violate the external constraint and join the coalition.

To summarise, given all subjects' inequality aversion is not strong enough, both the internal and external constraint are held. There exists a unique stable n^*-member coalition as we yield in Proposition 2. If the internal constraint is held, but the external constraint is violated, there exists a stable coalition which the size is larger than n^* members. If the internal constraint is violated, due to any subject having strong inequality aversion, there exists no coalition to be formed.

■

Appendix 1.3

Instructions

Please read the following instructions carefully.

You will have the guaranteed show-up fee £3. On top of that, you may – depending on your decisions and the decisions of others – earn more. There are three Parts in this experiment; in each Part there are several Rounds. The payoffs in each Round are independent: which means that the payoff in any one Round does not affect your payoffs in the following Rounds. At the end of each Part, a particular Round will be randomly selected and that will determine your payoff from that Part. Your total payment for this experiment is the sum-up your payoffs from these 3 Parts, plus a possible payoff from your partner in Parts 1 and 2. You will be paid in cash at the end of the experiment.

These Instructions are for your information. All subjects have identical Instructions. The experiment is anonymous. Please do not communicate with other participants during the experiment. If you have any questions, please let the experimenter know and he will answer you privately. We fear that if you violate this rule, we will have to exclude you from further participation in the experiment.

Part 1

Before starting the experiment, please answer the following questions:

- your user number, which is on the top of your monitor
- your major (Business, Economics, Humanities, Science, Laws, Engineering, Psychology, Others, pick up the one you belong to)
- the year you were born (in 4-digit format, e.g. 1980)
- your ethnicity (White, Mixed/multiple ethnic group, Asian/Asian British, Black/African/Caribbean/Black British, Other ethnic group)
- what level do you consider yourself as a religionist? (from 0 is *no religion* to 5 is *religionist*)
- what is your political preference? (0 is *left*, 1 is *centre-left*, 3 is *neutral*, 4 is *centre-right* and 5 is *right*)

The information will be kept confidential and used only in this study.

After answering the questions above, please click the "Start" button on your screen to proceed to the next stage of the experiment.

Now, you are going to start Part 1 of the experiment. You are playing with another subject in this room, who will be called your 'partner' in this Part. Everyone's identity and decisions will be anonymous and confidential.

There will be 11 *rounds* in this Part, preceded by a **trial round** for you to familiarise yourself with the game. In each round you will be asked to take a simple decision. Your payoff for this Part will depend upon your decision in a randomly chosen one of the 11 real rounds.

You are given £5 to share with your partner. There are 2 options:

- In Option 1, you get £2.5 and the partner gets £2.5 for sure.
- In Option 2, with a certain probability (which is different in different rounds), you get £5 and your partner gets nothing; with the residual probability, you get nothing and your partner gets £5.

There is an example of the decision problem on the screen. In each round, you will be given 30 seconds to make your decision. Your decision will be counted as Option 2 if you do not take a decision in these 30 seconds. At the end of Part 1, one of the 11 real rounds will be randomly chosen to determine your payment and that of your partner. The money you get from both your and your partner's decisions in this round will be paid to you at the end of the experiment.

Control Questions

The following questions are designed to help your understanding of the experiment.

Q1) Does the decision in one Round affect the decision in another Round?

Q2) Does the partner know your decision?

Given Option 2 is that, you have £5 and your partner has nothing with the probability 30%, or you have nothing and your partner has £5 with the probability 70%.

Q3) How much would you get if you choose Option 1?

Q4) How much would you get if you choose Option 2?

Part 2

This is also a decision problem. Your partner is reshuffled. He or she may be different to your partner in Part 1. Your identity and decisions will remain anonymous and confidential. There will be 20 *rounds* in this Part, preceded by a **trial round** for you to familiarise yourself with the game. In each round you will be asked to take a simple decision. Your payoff for this Part will depend upon your decision in a randomly chosen one of the 20 real rounds.

You are given 1 'token' to share with your partner. There are 2 options for you to choose.

- In Option 1, you keep the token.
- In Option 2, you give the token to the partner.

The value of the token to you and to your partner may differ, and are different in different rounds. There is an example of the decision problem on the screen.

In each round, you will be given 30 seconds to make your decision. Your decision will be counted as Option 2 if you do not take a decision in these 30 seconds. At the end of Part 2, one of the 20 real rounds will be randomly chosen to determine your payment and that of your partner. The money you get from both your and your partner's decisions in this round will be paid to you at the end of the experiment.

Control Questions

The following questions are designed to help your understanding of the experiment.

Q1) Does the decision in one Round affect the decision in another Round?

Q2) Does the partner know your decision?

Given that the value of the token is 50p for you and £1 for your partner.

Q3) How much would you get if you choose Option 1?

Q4) How much would you get if you choose Option 2?

Part 3

Please enter your user number.

This Part is different from Parts 1 and 2, in that you are now in a Group with **4** other players in this room. Your identity and decisions will remain anonymous and confidential. You will be indicated as a particular player in the Group, such as 'Player 1', 'Player 2' and so on. You will remain in this role in the same Group for the whole Part 3. Your payoff depends on the combination of your and other 4 players' decisions.

Your payoff for this Part will depend upon your decision in a randomly chosen one of the 60 real rounds. The whole session will take about 50 minutes.

In each round of this Part, you and each of the other 4 players in your Group have simply to decide, simultaneously and independently, whether or not to *join a coalition* with the other players. If you decide to join, please click 'YES'. If you decide not to join, please click 'NO'. If 2 or more players in your Group decide to join a coalition, then a coalition is said to be formed. If no-one decides to join, or if only 1 decides to join, then a coalition is *not* formed. If a coalition is not formed, everyone gets nothing.

There follows a sample payoff table, in which 'IN' means that the player has chosen to join the coalition and 'OUT' means that they have not.

- For the Trial Round and Rounds 1 to 15, please read Table 1.
- For Rounds 16 to 30, please read Table 2.
- For Rounds 31 to 45, please read Table 3.
- For Rounds 46 to 60, please read Table 4.

No one will know the decisions of the other players in the Group until all have made their decisions. When all have done so, all will be told the payoffs and decisions of all the players in the Group. Your decision has to be made within 180 seconds; otherwise the system will count your decision as that of 'not joining'.

Control Questions

The following questions are designed to help your understanding of the experiment.

Q1) Does the decision in one Round affect the decision in another Round?

Q2) In any Round do you know who has decided to join your Coalition before you take your decision?

Suppose that you are Player 3 in the Sample Table below.

Sample Payoff Table									
PLAYER 1		PLAYER 2		PLAYER 3		PLAYER 4		PLAYER 5	
IN	5.25	IN	5.25	IN	5.25	OUT	9	OUT	6.75
IN	4.5	IN	4.5	OUT	11.25	IN	4.5	OUT	6.75
IN	0	OUT	0	IN	0	IN	0	OUT	0
OUT	0	IN	0	IN	0	IN	0	OUT	0
IN	1.5	IN	1.5	OUT	7.5	OUT	6	OUT	4.5
IN	0	OUT	0	OUT	0	OUT	0	IN	0
OUT	0	IN	0	OUT	0	OUT	0	IN	0
OUT	0	OUT	0	IN	0	OUT	0	IN	0
OUT	0	OUT	0	OUT	0	IN	0	IN	0

Q3) Given the payoff table, Players 1 and 2 decide to join, and Player 4 and 5 decide not to join. How much you get if you choose 'YES'? (Note: Given you are Player 3)

Q4) Given the payoff table, Players 4 decides to join, and Player 1, 2 and 5 decide not to join. How much you get if you choose 'NO'? (Note: Given you are Player 3)

Table 1. Payoff Table in Trial Round and Rounds 1 to 15									
PLAYER 1		PLAYER 2		PLAYER 3		PLAYER 4		PLAYER 5	
IN	6.75	IN	6.75	IN	6.75	IN	6.75	IN	6.75
IN	0	IN	0	IN	0	IN	0	OUT	0
IN	0	IN	0	IN	0	OUT	0	IN	0
IN	0	IN	0	OUT	0	IN	0	IN	0
IN	3.75	OUT	12	IN	3.75	IN	3.75	IN	3.75
OUT	9	IN	4.5	IN	4.5	IN	4.5	IN	4.5
IN	0	IN	0	IN	0	OUT	0	OUT	0
IN	0	IN	0	OUT	0	IN	0	OUT	0
IN	0	OUT	0	IN	0	IN	0	OUT	0
OUT	0	IN	0	IN	0	IN	0	OUT	0
IN	0	IN	0	OUT	0	OUT	0	IN	0
IN	0	OUT	0	IN	0	OUT	0	IN	0
OUT	0	IN	0	IN	0	OUT	0	IN	0
IN	0	OUT	0	OUT	0	IN	0	IN	0
OUT	0	IN	0	OUT	0	IN	0	IN	0
OUT	6.75	OUT	9	IN	1.5	IN	1.5	IN	1.5
IN	0	IN	0	OUT	0	OUT	0	OUT	0
IN	0	OUT	0	IN	0	OUT	0	OUT	0
OUT	0	IN	0	IN	0	OUT	0	OUT	0
IN	0	OUT	0	OUT	0	IN	0	OUT	0
OUT	0	IN	0	OUT	0	IN	0	OUT	0
OUT	0	OUT	0	IN	0	IN	0	OUT	0
IN	0	OUT	0	OUT	0	OUT	0	IN	0
OUT	0	IN	0	OUT	0	OUT	0	IN	0
OUT	0	OUT	0	IN	0	OUT	0	IN	0
OUT	0	OUT	0	OUT	0	IN	0	IN	0

Table 2. Payoff Table in Rounds 16 to 30									
PLAYER 1		PLAYER 2		PLAYER 3		PLAYER 4		PLAYER 5	
IN	12.75	IN	12.75	IN	12.75	IN	12.75	IN	12.75
IN	0	IN	0	IN	0	IN	0	OUT	0
IN	0	IN	0	IN	0	OUT	0	IN	0
IN	8.25	IN	8.25	OUT	18	IN	8.25	IN	8.25
IN	9.75	OUT	12	IN	9.75	IN	9.75	IN	9.75
OUT	15	IN	9	IN	9	IN	9	IN	9
IN	0	IN	0	IN	0	OUT	0	OUT	0
IN	0	IN	0	OUT	0	IN	0	OUT	0
IN	0	OUT	0	IN	0	IN	0	OUT	0
OUT	0	IN	0	IN	0	IN	0	OUT	0
IN	0	IN	0	OUT	0	OUT	0	IN	0
IN	0	OUT	0	IN	0	OUT	0	IN	0
OUT	0	IN	0	IN	0	OUT	0	IN	0
IN	5.25	OUT	9	OUT	13.5	IN	5.25	IN	5.25
OUT	11.25	IN	4.5	OUT	13.5	IN	4.5	IN	4.5
OUT	11.25	OUT	9	IN	6	IN	6	IN	6
IN	0	IN	0	OUT	0	OUT	0	OUT	0
IN	0	OUT	0	IN	0	OUT	0	OUT	0
OUT	0	IN	0	IN	0	OUT	0	OUT	0
IN	0	OUT	0	OUT	0	IN	0	OUT	0
OUT	0	IN	0	OUT	0	IN	0	OUT	0
OUT	0	OUT	0	IN	0	IN	0	OUT	0
IN	0	OUT	0	OUT	0	OUT	0	IN	0
OUT	0	IN	0	OUT	0	OUT	0	IN	0
OUT	0	OUT	0	IN	0	OUT	0	IN	0
OUT	7.5	OUT	6	OUT	9	IN	1.5	IN	1.5

Table 3. Payoff Table in Rounds 31 to 45									
PLAYER 1		PLAYER 2		PLAYER 3		PLAYER 4		PLAYER 5	
IN	12	IN	12	IN	12	IN	12	IN	12
IN	9	IN	9	IN	9	IN	9	OUT	12
IN	6	IN	6	IN	6	OUT	24	IN	6
IN	10.5	IN	10.5	OUT	6	IN	10.5	IN	10.5
IN	0	OUT	0	IN	0	IN	0	IN	0
OUT	0	IN	0	IN	0	IN	0	IN	0
IN	3	IN	3	IN	3	OUT	18	OUT	9
IN	7.5	IN	7.5	OUT	4.5	IN	7.5	OUT	9
IN	0	OUT	0	IN	0	IN	0	OUT	0
OUT	0	IN	0	IN	0	IN	0	OUT	0
IN	4.5	IN	4.5	OUT	4.5	OUT	18	IN	4.5
IN	0	OUT	0	IN	0	OUT	0	IN	0
OUT	0	IN	0	IN	0	OUT	0	IN	0
IN	0	OUT	0	OUT	0	IN	0	IN	0
OUT	0	IN	0	OUT	0	IN	0	IN	0
OUT	0	OUT	0	IN	0	IN	0	IN	0
IN	1.5	IN	1.5	OUT	3	OUT	12	OUT	6
IN	0	OUT	0	IN	0	OUT	0	OUT	0
OUT	0	IN	0	IN	0	OUT	0	OUT	0
IN	0	OUT	0	OUT	0	IN	0	OUT	0
OUT	0	IN	0	OUT	0	IN	0	OUT	0
OUT	0	OUT	0	IN	0	IN	0	OUT	0
IN	0	OUT	0	OUT	0	OUT	0	IN	0
OUT	0	IN	0	OUT	0	OUT	0	IN	0
OUT	0	OUT	0	IN	0	OUT	0	IN	0
OUT	0	OUT	0	OUT	0	IN	0	IN	0

Table 4. Payoff Table in Rounds 46 to 60									
PLAYER 1		PLAYER 2		PLAYER 3		PLAYER 4		PLAYER 5	
IN	3	IN	3	IN	3	IN	3	IN	3
IN	1.5	IN	1.5	IN	1.5	IN	1.5	OUT	6
IN	0	IN	0	IN	0	OUT	0	IN	0
IN	0	IN	0	OUT	0	IN	0	IN	0
IN	0	OUT	0	IN	0	IN	0	IN	0
OUT	0	IN	0	IN	0	IN	0	IN	0
IN	0	IN	0	IN	0	OUT	0	OUT	0
IN	0	IN	0	OUT	0	IN	0	OUT	0
IN	0	OUT	0	IN	0	IN	0	OUT	0
OUT	0	IN	0	IN	0	IN	0	OUT	0
IN	0	IN	0	OUT	0	OUT	0	IN	0
IN	0	OUT	0	IN	0	OUT	0	IN	0
OUT	0	IN	0	IN	0	OUT	0	IN	0
IN	0	OUT	0	OUT	0	IN	0	IN	0
OUT	0	IN	0	OUT	0	IN	0	IN	0
OUT	0	OUT	0	IN	0	IN	0	IN	0
IN	0	IN	0	OUT	0	OUT	0	OUT	0
IN	0	OUT	0	IN	0	OUT	0	OUT	0
OUT	0	IN	0	IN	0	OUT	0	OUT	0
IN	0	OUT	0	OUT	0	IN	0	OUT	0
OUT	0	IN	0	OUT	0	IN	0	OUT	0
OUT	0	OUT	0	IN	0	IN	0	OUT	0
IN	0	OUT	0	OUT	0	OUT	0	IN	0
OUT	0	IN	0	OUT	0	OUT	0	IN	0
OUT	0	OUT	0	IN	0	OUT	0	IN	0
OUT	0	OUT	0	OUT	0	IN	0	IN	0

Appendix 2.1

Proof. A stable coalition requires both internal signatories stably stay in the coalition and external nonsignatories stably stay away from the coalition. Given that n^* is the smallest profitable coalition, an *unique equilibrium condition* ensure it is the only profitable coalition as

$$\gamma_{n^*} > \sum_{j=n^*+1}^{N} \gamma_j$$

With this condition, any signatory leaves the coalition, it collapses and all countries have nothing. The *internal constraint* will be satisfied as

$$S^s_{n^*}(n^*) > S^{ns}_{n^*}(n^* - 1) \tag{3.37}$$

which implies that any signatory can not be replaced by the participation of all nonsignatories. With this condition and the internal constraint, a coalition with n^* members are ensured stably.

From (2.4) , we can rewrite (3.37) as

$$S^s_i(n^*) > S^{ns}_i(n^* - 1)$$
$$\left(\sum_{i=1}^{n^*} \gamma_i - 1\right) + \theta_i \left[(n^* - 1)\left(\sum_{i=1}^{n^*} \gamma_i - 1\right) + \sum_{j=n^*+1}^{N} \gamma_j n^*\right] > 0$$

Because the coalition is profitable and no signatory can be replaced, a coalition with at least $n*$ members is stable internally. The altruism level θ_i does not affect to this constraint.

On the other hand, the *external constraint* requires all nonsignatories to

staying away from the coalition. It can be presented as

$$S_N^{ns}(n^*) > S_N^s(n^*+1) \tag{3.38}$$

The constraint can be rewritten as

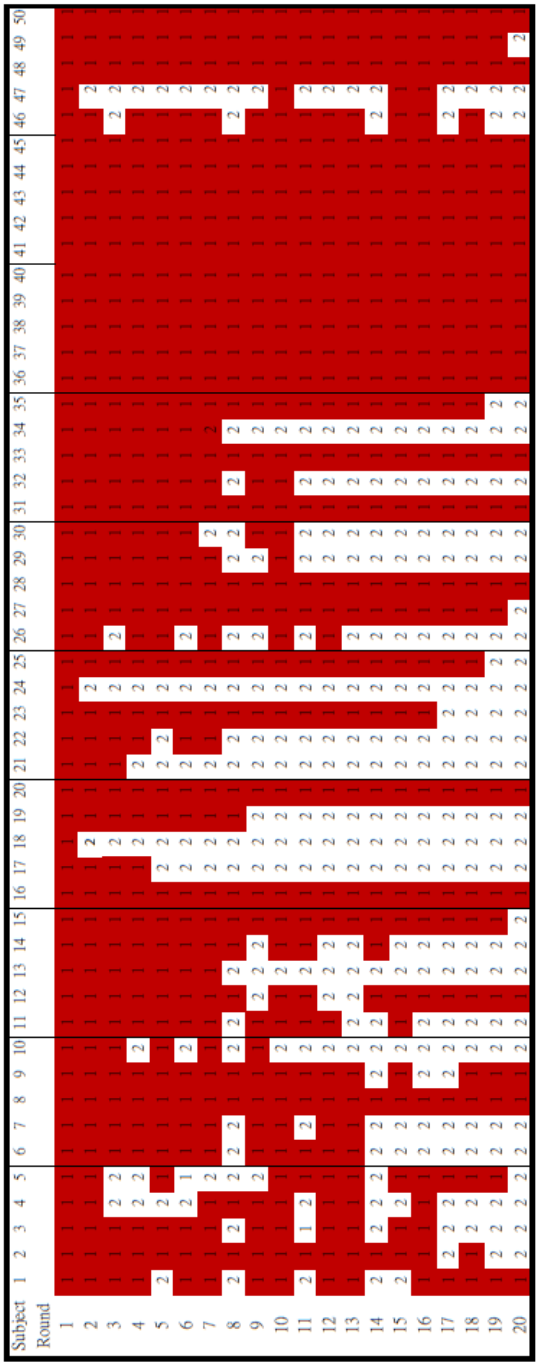
$$\begin{aligned} &\left(n^*\gamma_j\right) + \theta_j\left[n^*\left(\textstyle\sum_{i=1}^{n^*}\gamma_i - 1\right) + \textstyle\sum_j^{N-n^*-1}\left(n^*\gamma_j\right)\right] \\ &> \left(\textstyle\sum_{i=1}^{n}\gamma_i + \gamma_j - 1\right) + \theta_j\left[n^*\left(\textstyle\sum_{i=1}^{n^*}\gamma_i + \gamma_j - 1\right) + \textstyle\sum_j^{N-n^*-1}\left(n^*+1\right)\gamma_j\right] \end{aligned}$$

When a nonsignatory j is altruism neutral or its altruism level is not high enough ($\theta_j < \frac{\left[1+(n^*-1)\gamma_j - \sum_{i=1}^{n^*}\gamma_i\right]}{\left[n^*\gamma_j + \sum_{j'^*+1}^{N-1}\gamma_{j'}\right]}$), it would obey the external constraint and a n^*-member coalition is stable. Otherwise, j would violate the external constraint when it has high level of altruism. It implies that when j's altruism level (θ_j) is high, j is more likely to join the coalition and benefit to everyone. The size of the coalition would be bigger than n^*. Nevertheless, even some countries violate the external constraint and join the coalition, the internal constraint is ensured because of the unique equilibrium condition.

To summarise, the internal constraint can be satisfied with altruism but the external constraint may not be held. Hence, the size of the coalition is larger or equal to the smallest profitable coalition size n^*.

■

Appendix 2.2



Result of the Altruism Test

References

Andreoni, J. and Miller, J. (2002). "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica.* Vol. 70(2): 737-753.

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

Barbier, E. B. and Markandya, A. (1990). "The Conditions for Achieving Environmentally Sustainable Development". *European Economic Review.* Vol. 34(2-3): 659-669.

Barrett, S. (1994). "Self-Enforcing International Environmental Agreements." *Oxford Economic Paper.* Vol. 46(1): 878-894.

Barrett, S. (2001). "International Cooperation for Sale." *European Economic Review.* Vol. 45: 1835-1850.

Barrett, S. (2005). "Chapter 28. The Theory of International Environmental Agreements". In: Karl-Göran Mäler and Jeffrey R. Vincent, Editor(s), *Handbook of Environmental Economics.* Vol.3: 1457-1516.

Bettinger, E. and Slonim, R. (2006). "Using Experimental Economics to Measure the Effects of A Natural Educational Experiment on Altruism." *Journal of Public Economics*. Vol. 90(8-9): 1625-1648.

Blanco, M., Engelmann, D. and Normann, H.T. (2011). "A Within-Subject Analysis of Other-Regarding Preferences". *Games and Economic Behavior*. Vol. 72(2): 321-338.

Bohm, P. (2003). "Experimental Evaluations of Policy Instruments". In K. G. Mäler & J. R. Vincent (Eds.) *Handbook of environmental economics*. Vol. 1: 438–460.

Bratberg, E.; Tjotta, S.; and Oines, T. (2005). "Do Voluntary Internaiotnal Environmental Agreements Work?". *Journal of Environmental Economics and Management*. Vol. 50: 583-597.

Breton, M.; Sbragia, L.; and Zaccour, G. (2010). "A Dynamic Model for International Environmental Agreements". *Environmental and Resource Economics*. Vol. 45(1): 25-48.

Burger, N.E. and Kolstad, C.D. (2010). "International Environmental Agreements: Theory Meets Experimental Evidence" Working paper, University of California at Santa Barbara.

Carlsson, F.; Daruvala, D. and Johansson-Stenman, O. (2005). "Are People Inequality-Averse, or Just Risk-Averse?" *Economica*. Vol. 72(3): 375-396.

Carraro, C.; Marchiori, C. and Oreffice, S. (2009). "Endogenous Minimum Participation in International Environmental Treaties". *Environmental and Resource Economics*. Vol. 42(3): 411-425.

Carraro, C.; Siniscalco, D. (1998). "International Institutions and Environmental Policy: International environmental agreements: Incentives and political economy". *European Economic Review.* Vol. 42 (3–5): 561-572.

Charness, G. and Rabin, M. (2002). "Understanding Social Preferences With Simple Tests". *The Quarterly Journal of Economics.* Vol. 117(3): 817-869.

Dannenberg, A., Riechmann, T., Sturm, B. and Vogt, C. (2007). "Inequity Aversion and Individual Behavior in Public Good Games: An Experimental Investigation". Working paper.

D'Aspremont, C., Jacquemin, A., Gabszewicz, J., Weymark, J. (1983). "On the Stability of Collusive Price Leadership". *Canadian Journal of Economics.* Vol. 16(1): 17-25.

Dellink, R. and Finus, M. (2012). "Uncertainty and Climate Treaties: Does Ignorance Pay?" *Resource and Energy Economics.* Vol. 34(4): 565-584.

Eckel, C. and Lutz, N. (2003). "Introduction: What Role Can Experiments Play in Research on Regulation?". *Journal of Regulatory Economics*, Springer. Vol. 23(2): 103-07.

Eyckmans, J. and Finus, M. (2009). "An Almost Ideal Sharing Scheme for Coalition Games with Externalities." Discussion Paper. UK: University of Stirling.

Fehr, E. and Schmidt, K. M. (1999). "A Theory Of Fairness, Competition, And Cooperation". *The Quarterly Journal of Economics.* Vol. 114(3): 817-868.

Fischbacher, U., Gächter, S., Fehr, E., (2001). "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment". *Economics Letters*. Vol. 71(3): 397-404.

Fischbacher, U. (2007). "z-Tree: Zutich Toolbox for Ready-made Economic Experiments." *Experimental Economics*. Vol. 10(2): 171-178.

Finus, M., van Ierland, E. and Dellink, R. (2006). "Stability of Climate Coalitions in a Cartel Formation Game". *Economics of Governance*. Vol. 7(3): 271-291.

Finus, M. (2008). "Game Theoretic Research on the Design of International Environmental Agreements: Insights, Critical Remarks, and Future Challenges." *International Review of Environmental and Resource Economics* Vol. 2: 29–67.

Finus, M., Elena Saiz, M. and Hendrix, E.M.T. (2009). "An empirical test of new developments in coalition theory for the design of international environmental agreements". *Environment and Development Economics*. Vol. 14: 117-137.

Finus, M. and Pintassilgo, P. (2013). "The Role of Uncertainty and Learning for the Success of International Climate Agreements". *Journal of Public Economics*. Vol. 103: 29-43.

Georgescu-Roegen, N. (1971). *The Entropy Law and the Economic Process*. Cambridge, MA: Harvard University Press

Germain, M., Toint, P., Tulkens,H. and de Zeeuw, A. (2003). "Transfers to Sustain Dynamic Core-Theoretic Cooperation in International Stock Pollutant Control", *Journal of Economic Dynamics and Control*, Vol. 28(1): 79-99.

Greiner, B. (2004). "The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics." Working Paper Series in Economics 10, University of Cologne, Department of Economics.

Gruning, C. and Peters, W. (2010). "Can Justice and Fairness Enlarge International Environmental Agreements?". *Games*. Vol. 1(2): 137-158.

Hadjiyiannis, C.,D İriş, D., Tabakis, C. (2012). "International Environmental Cooperation under Fairness and Reciprocity." *The B.E. Journal of Economic Analysis and Policy*. Vol. 12(1): 1-30.

Hadjiyiannis, C., İriş, D. and Tabakis, C. (2012). "Multilateral Tariff Cooperation under Fairness and Reciprocity". *Canadian Journal of Economics/Revue canadienne d'économique.* Vol. 45: 925-941.

Hahn, R. W. and Ritz, R. A. (2014) "Optimal Altruism in Public-good Provision : An Application to Climate Policy". Working paper.

Hamilton, K. and Clemens, M. (1999). "Genuine Savings Rates in Developing Countries". *The World Bank Economic Review.* Vol. 13(2): 333-356.

Hartwick, J. M. (1977). "Intergenerational Equity and the Investing of Rents from Exhaustible Resources". *American Economic Review* Vol. 67(5): 972-74.

Helm, C. (1998). "International Cooperation Behind the Veil of Uncertainty – The Case of Transboundary Acidification." *Environmental and Resource Economics*. Vol. 12(2): 185-201.

Hoel, M. and Schneider, K. (1997). "Incentives to Participate in an International Environmental Agreement". *Environmental and Resource Economics.* Vol 9:153-170.

House of Lords (2005). *The Economics of Climate Change. Select Committee on Economic Affairs 2nd Report of Session 2005-06.* London. HL Paper 12-I.

Hwang, S.H., Bowles, S. (2012). "Is Altruism Bad for Cooperation?". *Journal of Economic Behavior and Organization.* Vol. 83(3): 330-341.

Karp, L. and Zhao, J. (2010). "International Environmental Agreements: Emissions Trade, Safety Valves and Escape Clauses". *Revue économique.* Vol. 61(1): 153-182.

Karp, L. (2013). "Provision of a Public good with Altruistic Overlapping Generations and Many Tribes". Working paper.

Kolstad, C. D. (2007). "Systematic Uncertainty in Self-Enforcing International Environmental Agreements". *Journal of Environmental Economics and Management.* Vol. 53(1): 68-79.

Kolstad, C. D. (2014) "Public Goods Agreements with Other-Regarding Preferences". Working paper

Kolstad, C. D. and Ulph, A. (2008). "Learning and International Environmental Agreements." *Climatic Change.* Vol. 89(1,2): 125-141.

Lessmann, K.; Marschinski, R. and Edenhofer, O. (2009). "The Effects of Tariffs on Coalition Formation in a Dynamic Global Warming Game". *Economic Modelling.* Vol. 26: 641-649.

Kroll, Y. and Davidovitz, L. (2003). "Inequality Aversion versus Risk Aversion". *Economica*. Vol. 70 (277): 19-29.

Kosfeld, M., Okada, A. and Riedl, A. (2009). "Institution Formation in Public Goods Games." *American Economic Review*. Vol. 99(4): 1335–55.

Kyoto Protocol. (1997). Kyoto Protocol to the United Nations Framework Convention on Climate Change.

Manne, A., Mendelsohn, R., and Richels, R. (1995). "MERGE: A Model for Evaluating Regional and Global Effects of GHG Reduction Policies". *Energy Policy*. Vol 23(1): 17-34.

Martinet, V. (2011). "A Characterization of Sustainability with Indicators". *Journal of Environmental Economics and Management*. Vol 61(2): 183-197.

Matros, A. (2012). "Altruistic Versus Egoistic Behavior in a Public Good Game". *Journal of Economic Dynamics and Control*, Vol. 36(4): 642-656.

Meadows, D. H., Meadows, D. L., Randers, J. and Behrens, W. W. (1972). *The Limits to Growth*.

McKibbin, W. J., Ross, M., Shackleton, R., and Wilcoxen, P. J. (1999). "Emissions Trading, Capital Flows and the Kyoto Protocol." *Energy Journal* (special issue).

Mitchell, R. B. (2003). "International Environmental Agreements Database." http://darkwing.uoregon.edu/~rmitchel/iea/

Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Clarendon Press.

Nagashima, M., Dellink, R., van Ierland, E., and Weikard, H.-P. (2009). "Stability of International Climate Coalitions — A Comparison of Transfer Schemes". *Ecological Economics*. Vol. 68(5): 1476-1487.

Nordhaus, W. D. (1994). *Managing the Global Commons: the Economics of Climate Change*. MIT Press, Cambridge.

Pearce, D.W., Markandya, A., and Barbier, E. (1989). *Blueprint for a Green Economy*, Earthscan: London

Pevnitskaya, S., and Ryvkin, D. (2011). "Behavior in a Dynamic Environment with Costs of Climate Change and Heterogeneous Technologies: An Experiment". in R. Mark Isaac, Douglas A. Norton (ed.) *Experiments on Energy, the Environment, and Sustainability. Research in Experimental Economics*. Vol. 14: 115-150.

Pezzey, J. C. V. (1997). "Sustainability constraints versus 'optimality' versus intertemporal concern, and axioms versus data". *Land Economics*, Vol 73(4): 448-466.

Riley, J. G. (1980). "The Just Rate of Depletion of a Natural Resource". *Journal of Environmental Economics and Management*. Vol. 7(4):291-307.

Rubio, S. J. and Ulph, A. (2007). "An Infinite-Horizon Model of Dynamic Membership of International Environmental Agreements". *Journal of Environmental Economics and Management*. Vol. 54(3): 296-310.

Solow, R. M. (1974). "Intergenerational Equity and Exhaustible Resources". *Review of Economic Studies*. pp. 29-46.

Stiglitz, J. (1974). "Growth with Exhaustible Natural Resources: Efficient and Optimal Growth Paths". *The Review of Economic Studies*. Vol. 41:123-137.

Toman, M.A. (1994). "Economics and "Sustainability": Balancing Trade-Offs and Imperatives". *Land Economics.* Vol. 70(4):399-413.

UNFCCC. (1992). FCCC/INFORMAL/84 GE.05-62220 (E) 200705 *http://unfccc.int/resource/docs/convkp/conveng.pdf*

van der Pol, T„ Weikard, H.-P., and van Ierland, E. (2012) "Can Altruism Stabilise International Climate Agreements?". *Ecological Economics*. Vol. 81: 112-120.

WCED (World Commission on Environment and Development). (1987). *Our Common Future*. Oxford, U.K.: Oxford University Press.

Weikard, H.P., Finus, M., and Altamirano-Cabrera, J.C. (2006). "The Impact of Surplus Sharing on the Stability of International Climate Agreements". *Oxford Economic Papers*. Vol. 58, 209-232.

Willinger, M. and Ziegelmeyer, A. (2001). "Strength of the Social Dilemma in A Public Goods Experiment: An Exploration of the Error Hypothesis". *Experimental Economics*. Vol. 4(2): 131-144.

Woodward, R. T. (2000). "Sustainability as Intergenerational Fairness: Efficiency, Uncertainty, and Numerical Methods". *American Journal of Agricultural Economics*. Vol. 82(3):581-593.

Yang, Y., Onderstal, S., and Schram, A. (2012). "Inequity Aversion Revisited". Working paper.

Yi, S.-S. (1997). "Stable Coalition Structures with Externalities". *Games and Economic Behavior.* Vol. 20(2): 201-237

de Zeeuw, A. (2008). "Dynamic Effects on the Stability of International Environmental Agreements". *Journal of Environmental Economics and Management.* Vol. 55(2): 163-174.