Environment for Development

Discussion Paper Series

October 2017 ■ EfD DP 17-12

Self-Enforcing International Environmental Agreements

The Role of Climate Tipping

Xin Liu, Lei Zhu, Xiao-Bing Zhang, and Magnus Hennlock

Environment for Development Centers

Central America
Research Program in Economics and Environment for Development in Central America Tropical Agricultural Research and Higher Education Center (CATIE)

Chile
Research Nucleus on Environmental and Natural Resource Economics (NENRE)
Universidad de Concepción

China
Environmental Economics Program in China (EEPC)
Peking University

Colombia
The Research Group on Environmental, Natural Resource and Applied Economics Studies (REES-CEDE), Universidad de los Andes, Colombia

Ethiopia
Environment and Climate Research Center (ECRC)
Ethiopian Development Research Institute (EDRI)

India
Centre for Research on the Economics of Climate, Food, Energy, and Environment, (CECFEE), at Indian Statistical Institute, New Delhi, India

Kenya
School of Economics
University of Nairobi

South Africa
Environmental Economics Policy Research Unit (EPRU)
University of Cape Town

Sweden
Environmental Economics Unit
University of Gothenburg

Tanzania
Environment for Development Tanzania
University of Dar es Salaam

USA (Washington, DC)
Resources for the Future (RFF)

Vietnam
University of Economics
Ho Chi Minh City, Vietnam

The Environment for Development (EfD) initiative is an environmental economics program focused on international research collaboration, policy advice, and academic training. Financial support is provided by the Swedish International Development Cooperation Agency (Sida). Learn more at www.efdinitiative.org or contact info@efdinitiative.org.

Self-Enforcing International Environmental Agreements: The Role of Climate Tipping

Xin Liu, Lei Zhu, Xiao-Bing Zhang, and Magnus Hennlock

Abstract

International environmental agreements (IEAs) are considered an important way to increase the efficiency of emission abatement and climate change mitigation. This paper uses a game-theoretic model to investigate the effect of possible tipping events, which would bring catastrophic and irreversible damage to ecological systems and human societies, on individual countries' emission decisions and on the scale of participation in IEAs that would maintain stability of the IEA. The results show that if the threshold of total emissions for triggering the tipping events is high, the possibility of climate tipping would have no effect on the stable number of participants in IEAs. However, if the threshold for tipping is low, whether the climate system would cross the tipping point depends on the magnitude of damage due to climate tipping. Moreover, the effect of possible tipping events would increase the size of a stable coalition if the signatories' coalition is the first mover. Otherwise, in a Nash equilibrium, the possibility of tipping would decrease the scope of participation in IEAs if the tipping damage is high, while it would induce more countries to sign IEAs if the tipping damage is moderate.

Key Words: international environmental agreements, climate change, catastrophic damage, tipping events

JEL Codes: C61, C73, D90, Q54

Contents

Self-Enforcing International Environmental Agreements: The Role of Climate Tipping

Xin Liu, Lei Zhu, Xiao-Bing Zhang, and Magnus Hennlock*

1. Introduction

When the rise in global temperature reaches a certain threshold, it is likely that there will be abrupt and irreversible events that can cause catastrophic damage to ecological systems and human societies, which are called tipping events. The fifth assessment report by the IPCC (Intergovernmental Panel on Climate Change) predicts catastrophes such as melting glaciers, sea level rise, loss of biodiversity, and so on, once the increase in temperature compared to the pre-industrial period exceeds 2℃(IPCC 2014a). The threshold for triggering these events is known as the ‘tipping point’, which has drawn the attention of political circles and the academic community in recent years (Russill and Nyssa 2009). Whether and when the tipping point will occur depends on how we regulate emissions. Therefore, climate policy could be reshaped as a response to the threat of climate tipping events.

It is clear that one country’s actions will influence other countries, because climate change is essentially a ‘public bad.’ The international community has agreed on the need for joint action to limit GHG emissions (IPCC 2014b). Cooperation among countries in reducing GHG emissions could generally improve efficiency in emission reductions, due to a shared decision-making objective and concerted action among players (Barrett 1994; McGinty 2006; Rubio and Ulph 2007). In order to promote such cooperation, the United Nations Framework Convention on Climate Change (UNFCCC) has conducted several rounds of discussions, leading to the signing of several remarkable agreements, e.g., the Kyoto Protocol and the Paris Agreement.

The threat of climate tipping events could have an influence on international cooperation, in that the possibility of these abrupt and irreversible events could affect each country’s decision on its emission strategies. Therefore, it is of great significance to investigate how the emission strategies of individual countries might change under the threat of climate tipping events and how this would affect their interest in participating in

*Xin Liu, Chinese Academy of Sciences. Lei Zhu, Beihang University. Xiao-Bing Zhang (corresponding author: xbzhmail@gmail.com), Renmin University of China, Beijing, PRC. Magnus Hennlock, IVL Swedish Environmental Research Institute.

international environmental agreements (IEAs) on climate change. In this paper, the possible tipping of climate is incorporated into a standard model of IEAs to investigate its effect on GHG emissions and international cooperation. Specifically, we build a game-theoretical model, which integrates the standard IEA framework with the possibility of climate tipping, to illustrate how the threat of climate tipping would affect individual countries' optimal emissions, their decisions about signing IEAs, and whether the climate system will cross the tipping point in equilibrium. We also investigate the effect of the tipping threat on the size of stable IEAs in a Stackelberg equilibrium, where the signatories' coalition is the first mover.

Our analysis reveals some interesting findings regarding the linkage between climate tipping and participation in IEAs. 1) If the threshold of total emissions for triggering the tipping events is high, the possibility of climate tipping would have no effect on the size of IEAs. 2) If the threshold for tipping is low, whether the climate system will cross the tipping point depends on the magnitude of damage due to climate tipping. 3) In the Nash equilibrium, the threat of climate tipping would decrease the size of IEAs if the tipping damage is high, while it would induce countries to sign IEAs if the tipping damage is moderate. 4) The threat of possible tipping events would increase the size of the coalition if the signatories' coalition is the first mover.

The remainder of the paper is organized as follows. Section 2 reviews the related studies in the literature. Section 3 presents the game and its equilibria in different cases. Section 4 summarizes the findings on the effect of the threat from climate tipping on the size of a stable IEA. Section 5 illustrates theoretical results with numerical simulations. Section 6 concludes.

2. Literature Review

International environmental agreements (IEAs) have been considered an important measure to improve cooperation on environmental protection and have been applied to various transboundary pollution issues, including ozone layer depletion, climate change, and so on (Finus 2002; Chander and Tulkens 1995). Barrett (1994) investigated the properties of self-enforcing IEAs and found that the number of signatories is small in coalitions that are both internally and externally stable. Barrett's work has been used and extended by many others with various perspectives and different focuses:

a) On the impacts of heterogeneous countries. For instance, Barrett (1995) divided countries into two groups and found that the conclusions derived from homogeneous countries can still hold. Fuentes and Rubio (2010) investigated the effect of payment transfers and heterogeneous abatement costs/benefits, showing that the assumption of heterogeneity had no significant impact on the size of the coalition, compared with the homogeneous case, if the only difference among countries was the abatement costs or if there were no transfers. With a large number of asymmetric nations and tradable pollution permits, McGinty (2006) found that the asymmetric case was similar to the symmetric case in the sense that the gains from cooperation decreased with the size of the coalition.

b) On the dynamic aspect of IEAs. For instance, Germain et al. (2003) found that the stability of a coalition needs to be ensured by transfers when the stock of pollution changes with time. Rubio and Ulph (2007) studied how the membership of an IEA would change with time, and found that the size of the IEA falls toward its steady state, while the stock of pollution rises toward its steady state. Breton et al. (2010) investigated the dynamic path of reaching a stable IEA when the signatories could punish the non-signatories, and showed that the size of a stable coalition depends on the initial pollution stock and the initial size of the coalition. Ulph (2004) compared fixed and variable membership and found that there would be higher social welfare with variable membership.

c) On the impact of uncertainties on IEAs. For instance, Iida (1993) studied the effect of modeling uncertainties on the characteristics of payoffs and found that uncertainties would undermine cooperation. Systematic uncertainties decrease the scope of participation in IEAs and, in the case of uncertainty regarding damage, the effect of learning depends on the parameters (Kolstad 2007). Kolstad and Ulph (2008) found that learning would generally reduce the global welfare from forming an IEA due to strategic interactions between decision-makers.

d) On the robustness of IEA size with respect to the non-parametric/parametric approaches. For instance, Karp and Simon (2013) showed that the conventional conclusions on the stable size of IEAs are not robust but depend on the parametric forms of cost/benefit functions. In addition, they showed that reductions in marginal abatement costs in an international environmental game can increase equilibrium membership. Lessmann et al. (2015) numerically identified robust results concerning the incentives of different nations to commit themselves to a climate agreement and estimated the extent of greenhouse gas mitigation that can be achieved by stable agreements.

Notably, some scholars have paid attention to the decision order of signatories and non-signatories in modeling IEAs, which could have an influence on individual countries' payoffs and could further affect the size of a stable coalition. While some studies assume that signatories and non-signatories make decisions simultaneously, i.e., they play a Nash game (Carraro and Siniscalco 1993; De Cara and Rotillon 2003; Finus and Rundshagen 2001; Rubio and Casino 2001), other scholars believe that the signatories can make decisions prior to non-signatories, i.e., countries play a Stackelberg game (Barrett 1994; McGinty 2006; Diamantoudi and Sartzetakis 2006; Rubio and Ulph 2007; İriş and Tavoni 2016). In the real world, signatories would sign an agreement with a commitment of future emissions, and non-signatories would consequently take the signatories' strategy into account when making decisions. The previous studies show that the size of a stable IEA may be different when they act in a Stackelberg fashion. Rubio and Ulph (2002) compared the results for a Nash game and those for a Stackelberg game, where they claimed that a stable IEA consists of three countries at most in a Nash game, while the size of a stable coalition in a Stackelberg game depends on the marginal damage.

The possible damage due to emissions is one of the most important concerns in the negotiation of an environmental agreement or climate treaty. As mentioned above, the rise of global temperature can push the global environmental system toward a threshold where sudden, irreversible events would occur and lead to remarkable and persistent damage (so called 'tipping events'). Lenton et al. (2008) considered a number of tipping elements in the climate system that could reach such thresholds in this century. Quite a few scientists have highlighted the catastrophic damage to economies and ecosystems that could be caused by abrupt climatic changes, and have suggested that decision-makers take into account such threats in policy design (Alley et al. 2003; Lenton 2011; Overpeck and Cole 2006).

Climate tipping events have also drawn the attention of environmental economists in recent years, triggering a wave of studies on the impacts of possible tipping events on economies and societies. A number of studies investigate optimal climate policies, e.g., carbon taxes, under the threat of tipping events by incorporating climatic thresholds in integrated assessment models (IAM) (Brozovic and Schlenker 2011; de Zeeuw and Zemel 2012; Cai et al. 2013; Lemoine and Traeger 2014). The results are very different when tipping events are considered compared to when they are omitted. For instance, Lemoine and Traeger (2014) used a recursive numerical climate-economy model to endogenize the probability of tipping and to learn about the location of the temperature threshold, and found that a greater possibility of climate tipping in the relatively near future would induce governments to increase carbon taxes. Mclnerney et al. (2012) incorporated the climatic threshold, parametric uncertainties, structural uncertainties and learning into the DICE-07 model, where they found that increasing the near-term investment in emission reduction would be a good way to improve the robustness of climatic strategies. De Zeeuw and Zemel (2012) analyzed how to manage risk with possible abrupt events under a dynamic framework, which could be applied to study pollution control under the threat of regime shifts. Barrett (2013) analyzed the countries' abatement behavior when taking catastrophic climate change into consideration and obtained conditions which could sustain full participation in climate treaties with certain and uncertain catastrophe. Nkuiya (2015) examined the effect of potential shift in damage on countries' behavior and their welfare in a transboundary pollution game when countries act non-cooperatively. Van der Ploeg and de Zeeuw (2016) investigated how climate catastrophes affect carbon taxes and precautionary saving when the North and South cooperate and when they do not. When they cooperate, carbon taxes would converge for developing and developed regions; otherwise, carbon taxes would diverge and there would be a bit more precautionary saving. Bahel (2017) built a dynamic model to analyze the global pollution problem with a critical threshold among heterogeneous countries and found that emissions are proportional to the gap between threshold and current stock in both non-cooperation and cooperation cases. Zhang and Zhu (2017) used a dynamic game model to investigate the effect of climate tipping events on strategic interaction between carbon taxes and energy prices, and found that tipping events would increase the carbon tax, decrease the wellhead fuel price, and shift the consumer price upward.

Recently, several studies have recognized the linkage between climate tipping and the participation of IEAs; those studies mainly focused on the effect of uncertainty in climate tipping (regime shift). Barrett (2013) and Barrett and Dannenberg (2014) designed various laboratory experiments to simulate countries' emission decisions with the possibility of climate tipping, where they found different tendencies to cooperate under a certain versus uncertain threshold of tipping. Nkuiya et al. (2015) focused on the effect of endogenous uncertainty of a regime shift in environmental damage on the participation in an IEA in a two-period framework, where the probability of a regime shift increases in the first-period pollution stock. They found that endogenous uncertainty appears to lower the pollution stock of both periods and increase the size of self-enforcing coalitions in period 1, compared to the case where the probability of the damage shift is exogenous. Schmidt (2017) investigated how a climate tipping point with unknown threshold would affect cooperation among countries, and found that climate tipping could improve the prospects of cooperation and that cooperation would allow countries to allocate their abatement efforts efficiently over time, leading to additional welfare gains from cooperation.

Focusing on the effect of climate tipping with a given threshold, in this paper we investigate the interactions between the climate tipping threat and international cooperation in various cases. We obtain analytic results regarding the stable sizes of coalitions and whether the climate system would cross the tipping point in equilibrium, with different locations of the tipping point and magnitudes of tipping damage.

3. Model

With n identical countries (indexed by $i=1,2,\cdots,n$), every country will choose its emissions q_i. Total emissions are $Q=\sum_{i=1}^{n} q_i$. Here, the tipping point refers to a threshold measured by the concentration (total emissions) of CO_2 in the atmosphere, such that an abrupt, irreversible change in the climatic system will occur once it has been passed (IPCC 2014; Alley et al. 2003; Overpeck and Cole 2006). Assume $\bar{Z}$ represents the triggering level of the amount of total CO_2 emissions in the atmosphere above which the tipping events would occur. The damage resulting from CO_2 concentration is assumed to be $D_i = bQ$ for each country if climate tipping does not happen (i.e., if $Q \leq \bar{Z}$),[1] where b is the marginal damage due to emissions. A jump in the damage will occur and each country will suffer from the tipping events if the climate system crosses the tipping point. The damage from the tipping events (so-called 'tipping damage') for one country is set as L and therefore the total damage in the case of total emissions passing the threshold (i.e., if $Q > \bar{Z}$) is $D_i^{tp} = bQ + L$.

Meanwhile, countries receive benefits from emissions, which are represented in a quadratic form, $R_i = aq_i - \frac{1}{2}q_i^2$, as in Dockner and Long (1993), McGinty (2006), and Breton et al. (2010), where a is the marginal benefit of the first unit of emissions.

[1] The assumption of linear damage would facilitate the closed form of solutions in the model, as highlighted by previous studies such as Breton et al. (2010). In fact, linear damage is not uncommon in the literature of IEAs; see, e.g., Barrett (1994), Kolstad (2007), Kolstad and Ulph (2008), and McGinty (2010).

Therefore, the net benefit (payoff) of country i is $\pi_i = aq_i - \frac{1}{2}q_i^2 - bQ$ if the climate tipping does not occur and $\pi_i^{tp} = aq_i - \frac{1}{2}q_i^2 - bQ - L$ if the climate tipping happens. The total net benefit of all countries is thus defined as $\Pi = \sum_{i=1}^{n}\pi_i$ if $Q \le \overline{Z}$ and $\Pi = \sum_{i=1}^{n}\pi_i^{tp}$ if $Q > \overline{Z}$.

3.1. The Fully Cooperative Case

In this case, all countries cooperate fully and decide their total emissions (Q^c) jointly. Therefore, the optimization problem for the coalition to maximize the total joint net benefits can be expressed as below, where the superscript ‘ c ’ denotes full cooperation.

$$\max_{Q^c} \Pi^c \quad s.t. Q^c \ge 0 \tag{1}$$

$$\text{where } \Pi^c = \begin{cases} aQ^c - \frac{1}{2n}(Q^c)^2 - nbQ^c, & Q^c \le \overline{Z} \\ aQ^c - \frac{1}{2n}(Q^c)^2 - nbQ^c - nL, & Q^c > \overline{Z} \end{cases}$$

As can be seen in (1), the fully-cooperative case of the game is degraded to a simple optimization problem. The objective function is piecewise but continuous in Q^c for each of its intervals (total emissions passing the threshold or not). The optimal value of total benefits in each interval can be derived through the first-order conditions, and the optimal emission decision under the full cooperation case (Q^{c*}) can therefore be obtained by comparing the maximized values in the two intervals (and choosing the larger one). The optimal solution would be the fully-cooperative equilibrium, and the results are summarized in Proposition 1. Note that each individual country’s emissions (denoted as $q^{c*} = Q^{c*} / n$) and net benefits (denoted as $\pi^{c*} = \Pi^{c*} / n$) are the same due to the symmetric-players assumption.[2]

[2] The subscripts are ignored due to the assumption of symmetric players.

Proposition 1.

When all countries cooperate,

1) If $\bar{Z} \geq n(a-nb)$, the total emissions are $Q^{c*} = n(a-nb)$; each country's emissions are $q^{c*} = a-nb$; the corresponding net benefits of each country are $\pi^{c*} = \frac{1}{2}(a-nb)^2$; and the climate system would not cross the tipping point in equilibrium.

2) If $\bar{Z} < n(a-nb)$ & $L \geq \frac{1}{2}(a-nb-\frac{\bar{Z}}{n})^2$, the total emissions are $Q^{c*} = \bar{Z}$; each country's emissions are $q^{c*} = \frac{\bar{Z}}{n}$; the corresponding net benefits of each country are $\pi^{c*} = -\frac{1}{2}(\frac{\bar{Z}}{n})^2 + (a-nb)\frac{\bar{Z}}{n}$; and the climate system would not cross the tipping point in equilibrium.

3) If $\bar{Z} < n(a-nb)$ & $L < \frac{1}{2}(a-nb-\frac{\bar{Z}}{n})^2$, the total emissions are $Q^{c*} = n(a-nb)$; each country's emissions are $q^{c*} = a-nb$; the corresponding net benefits of each country are $\pi^{c*} = \frac{1}{2}(a-nb)^2 - L$; and the climate system would cross the tipping point in equilibrium.

Proof. See Appendix A1.

Proposition 1 states that, if the threshold of total emissions that would trigger climate tipping (i.e., $\bar{Z}$) is high, the climate system would not cross the tipping point under full cooperation. Otherwise, whether the climate system would cross the tipping point depends on the magnitude of tipping damage, for the following reasons. If the tipping damage is high, it is optimal for countries to make an effort in emission control so as to prevent the occurrence of climate tipping; if the tipping damage is low, the individual countries do not pay enough attention to the tipping point.

3.2. The Non-Cooperative Case

In this case, each country makes its decision to maximize its own net benefits. That is, the representative country chooses its own emissions, taking the emissions of the other countries as given. Therefore, the optimization problem for the representative country in the game can be expressed as below, where the superscript 'oc' denotes non-cooperation.

$$\max_{q_i^{oc}} \pi_i^{oc} \quad s.t. q_i^{oc} \geq 0 \quad i = 1, 2, \cdots, n \tag{2}$$

$$\text{where } \pi_i^{oc} = \begin{cases} aq_i^{oc} - \frac{1}{2}(q_i^{oc})^2 - b(q_i^{oc} + \sum_{j \neq i} q_j^{oc}), & Q^{oc} \leq \bar{Z} \\ aq_i^{oc} - \frac{1}{2}(q_i^{oc})^2 - b(q_i^{oc} + \sum_{j \neq i} q_j^{oc}) - L, & Q^{oc} > \bar{Z} \end{cases}.$$

Calculating the first-order conditions for optimization in each interval, the best response function of each country can be obtained. Solving them simultaneously, the equilibrium condition can be derived. The detailed calculation can be seen in Appendix A2. The results of non-cooperative equilibria are summarized in Proposition 2. Note that q^{oc*} and π^{oc*} denote the equilibrium emissions and the corresponding payoffs for a country, respectively.

Proposition 2.

When there is no cooperation among countries,

1) If $\bar{Z} \geq n(a-b)$, the optimal emissions of each country are $q^{oc*} = a - b$; the corresponding net benefits are $\pi^{oc*} = \frac{1}{2}(a-nb)^2 - \frac{1}{2}(n-1)^2 b^2$; and the climate system would not cross the tipping point in equilibrium.

2) If $\bar{Z} < n(a-b)$ & $L \geq \frac{1}{2}(a-b-\frac{\bar{Z}}{n})^2$, the optimal emissions of each country are $q^{oc*} = \frac{\bar{Z}}{n}$; the corresponding net benefits are $\pi^{oc*} = -\frac{1}{2}(\frac{\bar{Z}}{n})^2 + (a-nb)\frac{\bar{Z}}{n}$; and the climate system would not cross the tipping point in equilibrium;

3) If $\bar{Z} < n(a-b)$ & $L < \frac{1}{2}(a-b-\frac{\bar{Z}}{n})^2$, the optimal emissions of each country are $q^{oc*} = a - b$;the corresponding net benefits are $\pi^{oc*} = \frac{1}{2}(a-nb)^2 - \frac{1}{2}(n-1)^2 b^2 - L$; and the climate system would cross the tipping point in equilibrium.

Proof. See Appendix A2.

It can be seen from the comparison of Propositions 1 and 2 that the emissions in the fully-cooperative equilibrium are generally smaller than those in the non-cooperative equilibrium. Table 1 summarizes the comparison of the fully-cooperative case and the non-cooperative case in terms of the emissions and net benefits of a representative country, where $\Delta q = q^{oc*} - q^{c*}$ is the difference in emissions and $\Delta \pi = \pi^{oc*} - \pi^{c*}$ is the difference in net benefits.

Table 1. Differences between the Fully Cooperative Case (‘c’) and the Non-Cooperative Case (‘oc’)

		Δq	$\Delta \pi$	Cross tipping point or not (‘c’|‘oc’)
$\bar{Z} \ge n(a-b)$		$(n-1)b$	$-\frac{1}{2}(n-1)^2 b^2$	NO|NO
	$L \ge \frac{1}{2}(a-b-\frac{\bar{Z}}{n})^2$	$\frac{\bar{Z}}{n}-(a-nb)$	$-\frac{1}{2}(\frac{\bar{Z}}{n}-(a-nb))^2$	NO|NO
	$L < \frac{1}{2}(a-b-\frac{\bar{Z}}{n})^2$	$(n-1)b$	$-\frac{1}{2}(n-1)^2 b^2 - L$	NO|YES
	$L \ge \frac{1}{2}(a-b-\frac{\bar{Z}}{n})^2$	0	0	NO|NO
	$\frac{1}{2}(a-nb-\frac{\bar{Z}}{n})^2 \le L < \frac{1}{2}(a-b-\frac{\bar{Z}}{n})^2$	$(a-b)-\frac{\bar{Z}}{n}$	$\frac{1}{2}(a-nb-\frac{\bar{Z}}{n})^2 - \frac{1}{2}(n-1)^2 b^2 - L$	NO|YES
	$L < \frac{1}{2}(a-nb-\frac{\bar{Z}}{n})^2$	$(n-1)b$	$-\frac{1}{2}(n-1)^2 b^2$	YES|YES

It can be seen from Table 1 that, in general, the emissions are higher and the net benefits are lower in the non-cooperative case, compared with the cooperative case. Also, the climate system is more likely to cross the tipping point in the non-cooperative case, which implies that cooperation could reduce the possibility of climate tipping by making individual countries emit less when taking into account the externality of their emissions, since they would maximize the total joint benefits rather than individual benefits. The incremental benefits from cooperation ($-\Delta\pi$) would always be non-negative, regardless of the threshold for tipping or the magnitude of tipping damage. Moreover, the emissions reduction due to cooperation (Δq) would increase with greater marginal damage[3] but decrease with a greater marginal benefit of emissions in most cases.

3.3. The Partially-Cooperative Case

The fully-cooperative case and the non-cooperative case investigated above are two of the extremes. A more realistic case would be that some countries participate in cooperation while others do not, i.e., the cooperation is partial only.

[3] This result is consistent with the conclusion in Barrett (1994).

Assume that there are m countries signing the IEA to form a coalition (‘ m ’ is an integer that is not greater than ‘ n ’), and that the remaining $n-m$ countries do not cooperate (i.e., they act individually). The objective of the coalition is to maximize all signatories’ total net benefits, while for the non-signatory the objective is to maximize its own net benefits individually. Superscripts ‘ s ’ and ‘ f ’ are used to denote the variables for signatories and non-signatories respectively. The respective optimization problems of the signatories’ coalition and the non-signatories can therefore be expressed as (3) and (4).

$$\max_{Q^s} \Pi^s \quad s.t. Q^s \geq 0 \tag{3}$$

$$\text{where } \Pi^s(m) = \begin{cases} aQ^s - \frac{1}{2m}(Q^s)^2 - mb(Q^s + \sum_j q_j^f), & Q \leq \bar{Z} \\ aQ^s - \frac{1}{2m}(Q^s)^2 - mb(Q^s + \sum_j q_j^f) - mL, & Q > \bar{Z} \end{cases}.$$

$$\max_{q_j^f} \pi_j^f \quad s.t. q_j^f \geq 0 \quad j = m+1, \cdots, n \tag{4}$$

$$\text{where } \pi_j^f(m) = \begin{cases} aq_j^f - \frac{1}{2}(q_j^f)^2 - b(q_j^f + \sum_{k \neq j} q_k^f + Q^s), & Q \leq \bar{Z} \\ aq_j^f - \frac{1}{2}(q_j^f)^2 - b(q_j^f + \sum_{k \neq j} q_k^f + Q^s) - L, & Q > \bar{Z} \end{cases}.$$

Due to the symmetric-player assumption, the emissions and net benefits of a signatory country would be $q_i^s(m) = \frac{Q^s(m)}{m}$ and $\pi_i^s(m) = \frac{\Pi^s(m)}{m}$ respectively. The total emissions of all countries are therefore $Q^{pc} = Q^s + \sum_j q_j^f$, where ‘ pc ’ represents ‘partial cooperation.’

3.3.1. Nash Equilibrium

By obtaining the reaction functions of signatories and non-signatories, the Nash equilibrium can be solved (see details in Appendix A3) and summarized in Proposition 3, where $q_{Nash}^{s*}(m)$, $\pi_{Nash}^{s*}(m)$, $q_{Nash}^{f*}(m)$ and $\pi_{Nash}^{f*}(m)$ denote the Nash equilibrium emissions and the corresponding payoffs for a signatory country and a non-signatory country, respectively.

Proposition 3.

With m countries signing the IEA,

1) If $\bar{Z} \geq n(a-b) + mb - m^2 b$, the respective emissions for a signatory and a non-signatory in the Nash equilibrium are $q_{Nash}^{s*}(m) = a - mb$ and $q_{Nash}^{f*}(m) = a - b$, and the corresponding

net benefits are $\pi^{s*}_{Nash}(m)=\frac{1}{2}(a-mb)^2-(n-m)(a-b)b$ and

$\pi^{f*}_{Nash}(m)=\frac{1}{2}(a-mb)^2-(n-m)(a-b)b+\frac{1}{2}(m^2-1)b^2$ respectively. The climate system would not cross the tipping point in the Nash equilibrium.

2) If $\bar{Z}<n(a-b)+mb-m^2b$ & $L\geq\frac{1}{2}(a-b-\frac{\bar{Z}+m(m-1)b}{n})^2$, the respective emissions for a signatory and a non-signatory in the Nash equilibrium are $q^{s*}_{Nash}(m)=\frac{\bar{Z}+(m-n)(m-1)b}{n}$ and $q^{f*}_{Nash}(m)=\frac{\bar{Z}+m(m-1)b}{n}$, and the corresponding net benefits are

$\pi^{s*}_{Nash}(m)=\frac{1}{2}a^2-\frac{1}{2}(a-\frac{\bar{Z}}{n}+\frac{(n-m)(m-1)b}{n})^2-b\bar{Z}$ and

$\pi^{f*}_{Nash}(m)=\frac{1}{2}a^2-\frac{1}{2}(a-\frac{\bar{Z}}{n}-\frac{m(m-1)b}{n})^2-b\bar{Z}$, respectively. The climate system would not cross the tipping point in the Nash equilibrium.

3) If $\bar{Z}<n(a-b)+mb-m^2b$ & $L<\frac{1}{2}(a-b-\frac{\bar{Z}+m(m-1)b}{n})^2$, the respective emissions for a signatory and a non-signatory in the Nash equilibrium are $q^{s*}_{Nash}(m)=a-mb$ and $q^{f*}_{Nash}(m)=a-b$, and the corresponding net benefits are

$\pi^{s*}_{Nash}(m)=\frac{1}{2}(a-mb)^2-(n-m)(a-b)b-L$ and

$\pi^{f*}_{Nash}(m)=\frac{1}{2}(a-mb)^2-(n-m)(a-b)b-L+\frac{1}{2}(m^2-1)b^2$, respectively. The climate system would cross the tipping point in the Nash equilibrium.

Proof. See Appendix A3.

According to Proposition 3, the form of the results under partial cooperation are similar to those under full cooperation and non-cooperation. Because no country can be forced to sign an IEA and signatories to an IEA can always withdraw from the agreement, the IEAs must be self-enforcing. Therefore it is worth investigating the scale of participation in self-enforcing IEAs that are stable in equilibrium. Referring to Barrett (1994), this paper adopts the definition of stability in oligopoly models (d'Aspremont et al. 1983), including internal stability (i.e., no signatory would like to become a non-signatory) and external stability (no non-signatory wants to become a signatory). We limit our analysis to the case of only one coalition formed by the signatories, and investigate the stability of this unique coalition, as in Barrett (1994), not taking into consideration the stability of the coalition's structure where there exist several coalitions.

Given m countries signing the IEA, let $\pi^s(m)$ and $\pi^f(m)$ denote net benefits of a signatory and a non-signatory, respectively, Considering internal and external stability, a stable IEA needs to meet the following conditions:

$$\begin{cases} \pi^s(m) \ge \pi^f(m-1) \\ \pi^f(m) \ge \pi^s(m+1) \end{cases} \tag{5}$$

From (5), one can obtain the size of a stable coalition under various circumstances, as summarized in Proposition 4.

Proposition 4.

The number of signatory countries, m^*_{Nash}, which would support a stable coalition in the Nash equilibrium, is as follows:

1) If $\bar{Z} \ge na - nb$, $m^*_{Nash} = 2$ or 3;
2) If $na - nb - 2b \le \bar{Z} < na - nb$:
   a) If $L \ge \frac{1}{2}(a - b - \frac{\bar{Z}}{n})^2$, $m^*_{Nash} = 3$;
   b) If $L < \frac{1}{2}(a - b - \frac{\bar{Z}}{n})^2$, $m^*_{Nash} = 2$ or 3;
3) If $\bar{Z} < na - nb - 2b$:
   a) If $L \ge \frac{1}{2}(a - b - \frac{\bar{Z}}{n})^2$, there is no stable coalition;
   b) If $\frac{1}{2}(a - b - \frac{\bar{Z} + X(k)b + X(k)^2 b}{n})^2 \le L < \frac{1}{2}(a - b - \frac{\bar{Z}}{n})^2$,
   $m^*_{Nash} = 2$ or 3 or $\left[\frac{1}{2} + \sqrt{\frac{1}{4} + n(a - b - \sqrt{2L}) - \bar{Z}}\right] + 1$;
   c) If $L < \frac{1}{2}(a - b - \frac{\bar{Z} + X(k)b + X(k)^2 b}{n})^2$, $m^*_{Nash} = 2$ or 3;
   where $k = \min(\left[\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{na - nb - \bar{Z}}{b}}\right], n)$, $X(k) = \min(\tilde{X}(k), k-1)$ and $\tilde{X}$ satisfies $\frac{n - \tilde{X}}{n}(k^2 - k - \tilde{X}^2 + \tilde{X}) = \frac{1}{2}(\tilde{X} - 1)(\tilde{X} - 3)$.

Proof. See Appendix A4.

The results summarized in Proposition 4 have several implications:

1) If the threshold of total emissions for tipping is sufficiently high, or the threshold is low but the tipping damage is also low, there would be no large-scale stable cooperation. Coalitions with only two or three countries would be stable, which is in accordance with previous research on the absence of tipping events (Carraro and Siniscalco 1993; Hoel 1992).

2) If the threshold for tipping is low but the tipping damage is high, there would be no stable coalition, and the climate system would not cross the tipping point in the Nash equilibrium. A signatory's total net benefit on average is always smaller than that of a non-signatory (as can be seen in Proposition 3); this implies that the non-signatories could free-ride on the damage avoided by the signatories. The IEA is not stable because the signatories' coalition possesses only external stability but not internal stability. Therefore, there would be no stable coalition under this circumstance.

3) If the threshold of total emissions is low and the tipping damage is moderate, the largest size of a stable coalition is less than $\frac{2}{3}n$ but increases as the threshold for tipping decreases, which implies that more countries would sign an IEA if the current concentration level is closer to the threshold. In addition, the size of a stable coalition would decrease with greater tipping damage. Furthermore, under this circumstance, no signatory would leave the coalition because it would gain less (i.e., if a signatory left, the climate system would cross the tipping point), which implies that the internal stability of a coalition would be satisfied. Thus, the threat of climate tipping enlarges the size of a stable coalition under this circumstance.

3.3.2. Stackelberg Equilibrium

In the Nash equilibrium presented above, the signatories and non-signatories make their decisions simultaneously. In reality, however, it is possible that the signatories could make decisions in advance (first mover), since the signatories need to make promises about their emissions when signing IEAs, as highlighted in the literature, e.g., Rubio and Ulph (2002) and İriş and Tavoni (2016). Therefore, in what follows, we also present the Stackelberg equilibrium, where the signatories' coalition is treated as a leader that is the first mover (in choosing its emissions level), while the non-signatories are followers that decide their emissions levels based on the coalition's strategy. The Stackelberg equilibrium is summarized in Proposition 5， and the detailed solution can be found in Appendix A5. Note that $q_{Stack}^{s*}(m)$, $q_{Stack}^{f*}(m)$, $\pi_{Stack}^{s*}(m)$ and $\pi_{Stack}^{f*}(m)$ in Proposition 5 denote Stackelberg equilibrium emissions for a signatory and a non-signatory's corresponding net benefits, respectively. Also, $i = 1, 2, \cdots m$, $j = 1, 2, \cdots n - m$.

Proposition 5.

With m countries signing an IEA,

1) If $\overline{Z} \geq n(a-b) + mb - m^2 b$, the respective emissions for signatories and non-signatories in the Stackelberg equilibrium are $q_i^{s*}(m) = a - mb$ and $q_j^{f*}(m) = a - b$, and the corresponding net benefits are $\pi_i^{s*}(m) = \frac{1}{2}(a - mb)^2 - (n-m)(a-b)b$ and $\pi_j^{f*}(m) = \frac{1}{2}(a - mb)^2 - (n-m)(a-b)b + \frac{1}{2}(m^2 - 1)b^2$, respectively. The climate system would not cross the tipping point in the Stackelberg equilibrium.

2) If $\overline{Z} < n(a-b) + mb - m^2 b$ & $L \geq \frac{1}{2}(a - b - \frac{\overline{Z} - \min(ma, \overline{Z})}{n-m})^2$, the respective emissions for signatories and non-signatories in the Stackelberg equilibrium are $q_i^{s*}(m) = \min(a, \frac{\overline{Z}}{m})$ and $q_j^{f*}(m) = \frac{\overline{Z} - \min(ma, \overline{Z})}{n-m}$, and the corresponding net benefits are $\pi_i^{s*}(m) = \min(\frac{1}{2}a^2 - b\overline{Z}, a\frac{\overline{Z}}{m} - \frac{1}{2}(\frac{\overline{Z}}{m})^2 - b\overline{Z})$ and $\pi_j^{f*}(m) = \max(\frac{1}{2}a^2 - \frac{1}{2}(\frac{na - \overline{Z}}{n-m})^2 - b\overline{Z}, -b\overline{Z})$, respectively. The climate system would not cross the tipping point in the Stackelberg equilibrium;

3) If $\overline{Z} < n(a-b) + mb - m^2 b$ & $\frac{1}{2}[A(m)]^2 \leq L < \frac{1}{2}(a - b - \frac{\overline{Z} - \min(ma, \overline{Z})}{n-m})^2$, the respective emissions for signatories and non-signatories in the Stackelberg equilibrium are $q_i^{s*}(m) = \frac{\overline{Z} - (n-m)(a-b)}{m} + \frac{n-m}{m}\sqrt{2L}$ and $q_j^{f*}(m) = a - b - \sqrt{2L}$, and the corresponding net benefits are $\pi_i^{s*}(m) = \frac{1}{2}a^2 - \frac{1}{2}(\frac{na - (n-m)b - \overline{Z}}{m} - \frac{n-m}{m}\sqrt{2L})^2 - b\overline{Z}$ and $\pi_j^{f*}(m) = \frac{1}{2}a^2 - \frac{1}{2}(b + \sqrt{2L})^2 - b\overline{Z}$, respectively. The climate system would not cross the tipping point in the Stackelberg equilibrium;

4) If $\overline{Z} < n(a-b) + mb - m^2 b$ & $L < \frac{1}{2}[A(m)]^2$, the respective emissions for signatories and non-signatories in the Stackelberg equilibrium are $q_i^{s*}(m) = a - mb$ and $q_j^{f*}(m) = a - b$, and the corresponding net benefits are $\pi_i^{s*}(m) = \frac{1}{2}(a - mb)^2 - (n-m)(a-b)b - L$ and $\pi_j^{f*}(m) = \frac{1}{2}(a - mb)^2 - (n-m)(a-b)b + \frac{1}{2}(m^2 - 1)b^2 - L$, respectively. The climate system

would cross the tipping point in the Stackelberg equilibrium, where

$A(m)=\dfrac{(n(a-b)+mb-m^2b-\bar{Z})^2}{2(n-m)(n(a-b)+mb-\bar{Z})}$ if $m=\dfrac{n}{2}$, and

$$A(m)=\frac{-(n(a-b)+mb-\bar{Z})(n-m)+\sqrt{(n(a-b)+mb-\bar{Z})^2(n-m)^2+(m^2-(n-m)^2)(n(a-b)+mb-m^2b-\bar{Z})^2}}{m^2-(n-m)^2}$$ if

$m\neq\dfrac{n}{2}$.

Proof. See Appendix A5.

By comparing Propositions 3 and 5, the following observations can be obtained: 1) if the threshold of total emissions is sufficiently high, or the threshold is low but the tipping damage is also low, the results of the Nash equilibrium and the Stackelberg equilibrium would be the same, because in this circumstance the reaction functions of countries are independent of others' action. 2) If the threshold of total emissions is low but the tipping damage is high, the signatories' benefits would increase while non-signatories' benefits would decrease in the Stackelberg equilibrium, compared with the case in the Nash equilibrium, which reflects the signatories' first-mover advantage.

In the Stackelberg game, if the threshold is low and the tipping damage is high, because the signatories' coalition tends to prevent the climate system from crossing the tipping point, its optimal strategy is to leave some emitting room ($\bar{Z}-Q^s$) for the non-signatories to assure that their gains are no less than those in the case of emitting more and triggering the tipping. Such emitting room is smaller than that in the Nash game case, which implies that the signatories may emit more and therefore non-signatories may emit less.

Proposition 6 summarizes how the stable size of the signatories' coalition would change with respect to the threshold of total emissions and tipping damage.

Proposition 6

The number of signatory countries m^*_{Stack} that would support a stable coalition in the Stackelberg game follows:

1) If $\bar{Z}\geq na-nb-6b$, $m^*_{Stack}=2$ or 3;

2) If $na-nb+\left[\dfrac{n}{2}\right]b-\left[\dfrac{n}{2}\right]^2 b\leq\bar{Z}<na-nb-6b$,

a) If $L > \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = \left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]$;

b) If $L = \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = 2,3,\cdots,\left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]$;

c) If $L < \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = 2$ or 3;

3) If $na-nb-\left[\frac{n-1}{2}\right]b-\left[\frac{n-1}{2}\right]^2 b \le \overline{Z} < na-nb+\left[\frac{n}{2}\right]b-\left[\frac{n}{2}\right]^2 b$

a) If $L \ge \frac{1}{2}(a-b-\frac{\overline{Z}-\min((k_S-1)a,\overline{Z})}{n-(k_S-1)})^2$, $m^*_{Stack} = \left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]+1$;

b) If $\frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2 < L < \frac{1}{2}(a-b-\frac{\overline{Z}-\min((k_S-1)a,\overline{Z})}{n-(k_S-1)})^2$,

$m^*_{Stack} = \left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]$;

c) If $L = \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = 2,3,\cdots,\left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]$;

d) If $L < \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = 2$ or 3;

4) If $na-nb-Ub-U^2b \le \overline{Z} < na-nb-\left[\frac{n-1}{2}\right]b-\left[\frac{n-1}{2}\right]^2 b$

a) If $L > \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = \left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]+1$;

b) If $L = \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = 2,3,\cdots,\left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]+1$;

c) If $\frac{1}{2}A(k_S-1)^2 \le L < \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = 2$ or 3 or $\left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]+1$;

d) If $L < \frac{1}{2}A(k_S-1)^2$, $m^*_{Stack} = 2$ or 3;

5) If $na-n^2b \le \overline{Z} < na-nb-Ub-U^2b$

a) If $L > \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = \left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]+1$;

b) If $L = \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = 2,3,\cdots,\left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]+1$;

c) If $\frac{1}{2}A(k_S)^2 \le L < \frac{1}{2}(a-b-\frac{\overline{Z}}{n})^2$, $m^*_{Stack} = 2$ or 3 or $\left[\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{na-nb-\overline{Z}}{b}}\right]+1$;

d) If $L < \frac{1}{2}A(k_S)^2$, $m^*_{Stack} = 2$ or 3;

6) If $\bar{Z} < na - n^2 b$

a) If $L > \frac{1}{2}(a - b - \frac{\bar{Z}}{n})^2$, $m^*_{Stack} = n$;

b) If $L = \frac{1}{2}(a - b - \frac{\bar{Z}}{n})^2$, $m^*_{Stack} = 2, 3, \cdots, n$;

c) If $L < \frac{1}{2}(a - b - \frac{\bar{Z}}{n})^2$, $m^*_{Stack} = 2$ or 3.

where $k_S = \left[\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{na - nb - \bar{Z}}{b}} \right]$, and U satisfies

$(U+1)^2 b^2 = (A(U) + b)^2 + 2b(\bar{Z} - na + nb + Ub + U^2 b)$ and $U > \left[\frac{n}{2} \right]$.

Here, $A(\cdot)$ is the same function as that in Proposition 5.

Proof. See Appendix A6.

From Proposition 6, it can be seen that, if the threshold of total emissions is low but the tipping damage is also low, a stable coalition would have only two or three countries, i.e., there is no large-scale stable cooperation. If the tipping damage is moderate, more countries tend to participate in the IEA as the threshold of total emissions decreases, which leads to a larger scale of cooperation. If the tipping damage is sufficiently high, all countries tend to be signatories. This is because the coalition's first-mover advantage would bring signatories more benefit than non-signatories, which implies that in this circumstance a coalition would have internal stability but not external stability, and thus only a grand coalition would be stable. This result is similar to Rubio and Ulph (2002)'s conclusion that a grand coalition is stable if marginal damage is sufficiently high. Compared with the result with no stable coalition in the Nash equilibrium, it can be seen that the first-mover advantage could induce more countries to join the coalition, which is consistent with the findings in Rubio and Ulph (2002) that the first-mover advantage could increase the size of a stable coalition and decrease free-riding behaviors.

In addition, the critical value of the tipping damage (i.e., $\frac{1}{2}A(k_S)^2$ in Proposition 6), which determines whether the size of a stable coalition is larger, increases as the threshold for tipping increases. This implies that, if the threshold for tipping is closer to the current concentration level, even low tipping damage could stimulate countries to cooperate. That is, when the situation of the climate system is threatening, countries would tend to cooperate to deal with it.

4. Discussion

In Section 3, we saw that countries' strategies and whether the climate system would cross the tipping point differ from one another in fully-cooperative, non-cooperative and partially-cooperative cases, and that the size of a stable coalition would be different in the Nash equilibrium and the Stackelberg equilibrium. In this section, we will discuss the effect of different factors on the size of a stable coalition and whether the climate system would cross the tipping point.

4.1. The Effect of the Climate Tipping Threat on the Size of a Stable Coalition

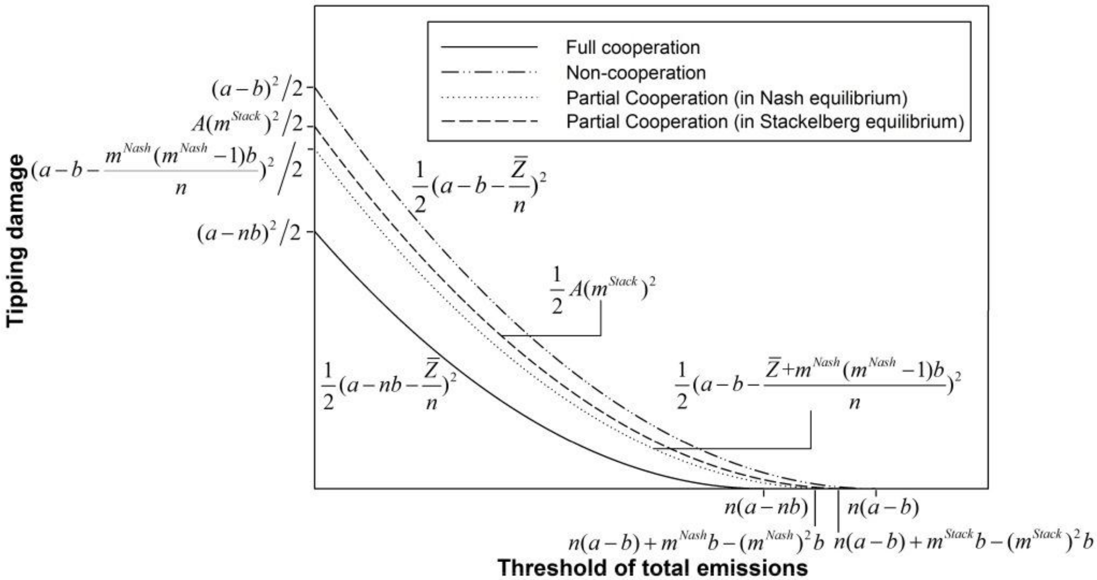
As shown above, if there is no threat of tipping in the climate system, which would be equivalent to the case with a sufficiently high threshold for tipping, only the coalitions with two or three countries are stable. That is, only a small number of countries would sign the IEA voluntarily.

If there exists the threat of climate tipping (i.e., the threshold for tipping is low), the size of a stable coalition would depend on the threshold for tipping, the magnitude of the tipping damage and whether the coalition is the first mover. Specifically, 1) if the tipping damage is also low, the size of a stable coalition is the same as that without the threat of climate tipping (i.e., only the coalitions with two or three countries are stable). 2) If the tipping damage is moderate, the threat of tipping would stimulate more countries (but less than $\frac{2}{3}n$) to cooperate, because a small coalition would not prevent the climate system from crossing the tipping point. 3) If the tipping damage is sufficiently high, there would be no stable coalition, i.e., the threat of climate tipping would decrease the size of a stable coalition. However, if the signatories' coalition can move first, the grand coalition would be stable, i.e., the tipping threat promotes the signing of IEAs under this circumstance.

4.2. The Factors Influencing Whether the Climate Tips

Figure 1 shows the boundaries of whether the climate system crosses tipping point with different thresholds and tipping damage in different cooperative cases. The climate system would cross the tipping point when the point is located in the area below the boundary, that is, the threshold for tipping and the tipping damage are both lower than certain critical values, which depend on the size of the coalition and parameters in the benefit and damage functions.

Figure 1. Summary of Areas Where the Climate System Would Tip under Various Circumstances



4.2.1. The Effect of the Coalition's Size

It can be seen from Fig. 1 that the critical values of the threshold and damage which determine whether the climate system would tip are functions of the size of the signatories' coalition. Assuming that the size of the coalition (i.e., m) is given, we can investigate the effect of the coalition's size on the outcome of the climate. 1) For the critical value of the threshold for tipping $\bar{Z}^{\#}(m) = -m^2 b + mb + n(a-b)$, the derivative with respect to the size of the coalition is $\frac{\partial \bar{Z}^{\#}(m)}{\partial m} = -2bm + b$. With $m > \frac{1}{2}$, we have $\frac{\partial \bar{Z}^{\#}(m)}{\partial m} < 0$. 2) For the critical value of the tipping damage $L^{\#}(m) = \frac{1}{2}(a-b-\frac{\bar{Z}+m(m-1)b}{n})^2$, the derivative with respect to the size of the coalition, $\frac{\partial L^{\#}(m)}{\partial m}$, would be negative because $m \geq 1$. That is, the critical values of the tipping threshold and the tipping damage decrease with the size of the coalition. Consequently, we know that the climate system is less likely to cross the tipping point with a larger size of coalition, because a larger size of coalition leads to less free-riding and less total emissions.

4.2.2. The Effect of Benefit and Damage Parameters

In what follows, we investigate the effect of parameter 'a' in the benefit function on whether the climate system would cross the tipping point, still assuming a given size of the coalition, m. The partial derivatives of the critical values of the tipping threshold and tipping damage with respect to 'a' are $\frac{\partial \bar{Z}^{\#}(a)}{\partial a} = n > 0$ and $\frac{\partial L^{\#}(a)}{\partial a} > 0$, respectively, implying that both critical values increase with 'a'. This indicates that the climate system is more likely to cross the tipping point with a larger 'a', i.e., with larger benefit from emissions.

Analogously, one can investigate the effect of the damage parameter 'b'. The partial derivatives of the critical values with respect to 'b' are $\frac{\partial \bar{Z}^{\#}(b)}{\partial b} < 0$ and $\frac{\partial L^{\#}(b)}{\partial b} < 0$, respectively. This implies that the climate system is less likely to cross the tipping point with an increase of 'b', due to the more aggressive emission reduction when the marginal damage is large.

5. Numerical Illustrations

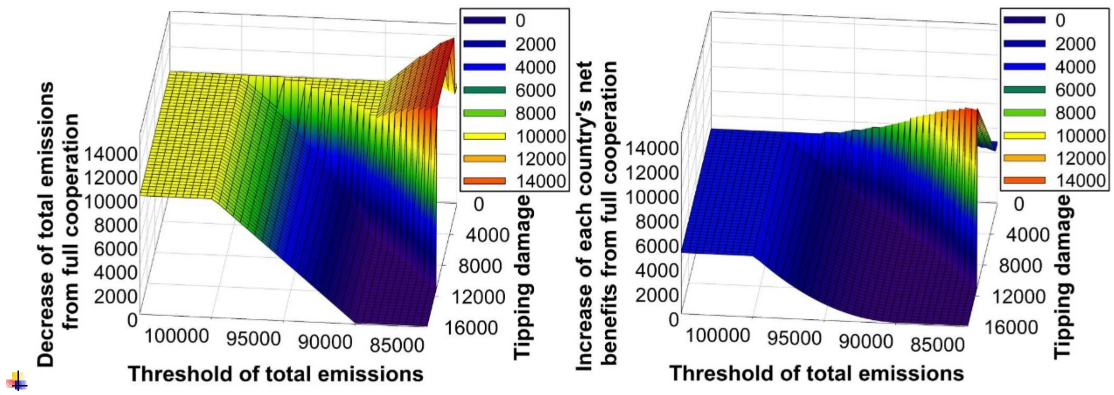
To complement the theoretical analysis above, in this section we conducted various numerical simulations to test the predictions of the game model. Referring to Barrett (1994), the parameters are a=1000, b=1, n=100. We let the threshold of total emissions vary from 85000 to 105000 and tipping damage vary from 0 to 16000, in order to examine how the results would change with the parameters[4]. Based on the propositions in Section 3, we can obtain the equilibrium strategy under different cases, and then determine whether a coalition with a certain size is stable based on the stability conditions mentioned above (see the inequalities in (5)).

5.1. Benefit from Full Cooperation

The numerical results in the fully-cooperative and non-cooperative cases are illustrated in Fig. 2, from which it can be seen that full cooperation leads to a decrease of total emissions and increase of each country's net benefits.

[4] We take a value of the threshold every 500; then, there are 41 values. Taking a value of tipping damage every 400, there are also 41 values. Thus, there are a total of 1681 parameter combinations of the threshold and tipping damage, i.e., $(\bar{Z}, L)$.

Figure 2. Effects of Cooperation on Emissions and Benefits



(a) (b)

When $\bar{Z}=85000$, $1600 \le L \le 10800$ – i.e., when the threshold for tipping is lowest and the tipping damage is moderate – full cooperation leads to the most dramatic emission reduction (14900). The largest incremental benefits from full cooperation (14451) occur when $\bar{Z}=85000$, $L=10800$. Due to the lower critical value of tipping damage in the fully cooperative case, when the tipping damage is moderate, in the fully cooperative case the optimal decision of countries is to make their total emissions just equal to the threshold to avoid the tipping damage. In the non-cooperative case, countries would emit more to maximize their own benefits, where the total emissions in equilibrium (which is a constant) are higher than the threshold. Under this circumstance, countries in the non-cooperative case would suffer tipping damage while countries in the fully cooperative case would not. Thus, the largest incremental benefits occur with the greatest tipping damage within this interval.

We can also see that, if the threshold for tipping is low but the tipping damage is high, there is no difference in emissions and benefit between these two cases. For instance, when $\bar{Z}=87500$, $L=12000$, the emission reduction and incremental benefits are both zero. This is because countries would emit less to avoid the great tipping damage in this situation, whether or not they are in cooperation.

5.2. The Size of a Stable IEA

Table 2 summarizes the size of a stable signatories' coalition under different parameter combinations. For instance, in both the Nash and the Stackelberg equilibrium, when $\bar{Z}=100000$, $L=16000$, and $\bar{Z}=85000$, $L=2000$, we have $\begin{cases}\pi^{s*}(2)>\pi^{f*}(1)\\ \pi^{f*}(2)=\pi^{s*}(3)\end{cases}$ and

$\begin{cases} \pi^{s*}(3) = \pi^{f*}(2) \\ \pi^{f*}(3) > \pi^{s*}(4) \end{cases}$, which means that the coalitions consisting of two or three countries are stable. When $\bar{Z} = 90000, L = 16000$, in the Nash equilibrium we have $\pi_{Nash}^{s*}(m) < \pi_{Nash}^{f*}(m-1)$ for all $m = 2, \cdots, n$, i.e., a signatory's net benefit is always less than that of a non-signatory, implying that there is no stable coalition; while in the Stackelberg equilibrium, $\pi_{Stack}^{s*}(m) > \pi_{Stack}^{f*}(m-1)$ for $m = 2, \cdots, n$ would hold, i.e., a signatory's net benefit is always larger than that of a non-signatory, implying that the coalition would be stable only when all countries sign the IEA. When $\bar{Z} = 87500$, $L = 6000$, in the Nash equilibrium we have $\begin{cases} \pi^{s*}(2) > \pi^{f*}(1) \\ \pi^{f*}(2) = \pi^{s*}(3) \end{cases}$, $\begin{cases} \pi^{s*}(3) = \pi^{f*}(2) \\ \pi^{f*}(3) > \pi^{s*}(4) \end{cases}$ and $\begin{cases} \pi^{s*}(39) > \pi^{f*}(38) \\ \pi^{f*}(39) > \pi^{s*}(40) \end{cases}$, which implies that the coalitions consisting of 2 or 3 or 39 countries are stable; while in the Stackelberg equilibrium, we have $\begin{cases} \pi^{s*}(2) > \pi^{f*}(1) \\ \pi^{f*}(2) = \pi^{s*}(3) \end{cases}$ and $\begin{cases} \pi^{s*}(3) = \pi^{f*}(2) \\ \pi^{f*}(3) > \pi^{s*}(4) \end{cases}$, which implies that the coalitions consisting of two or three countries are stable. This is consistent with the prediction of Propositions 4 and 6.

Table 2. The Stable Number of Participants in IEA under Different Parameter Combinations

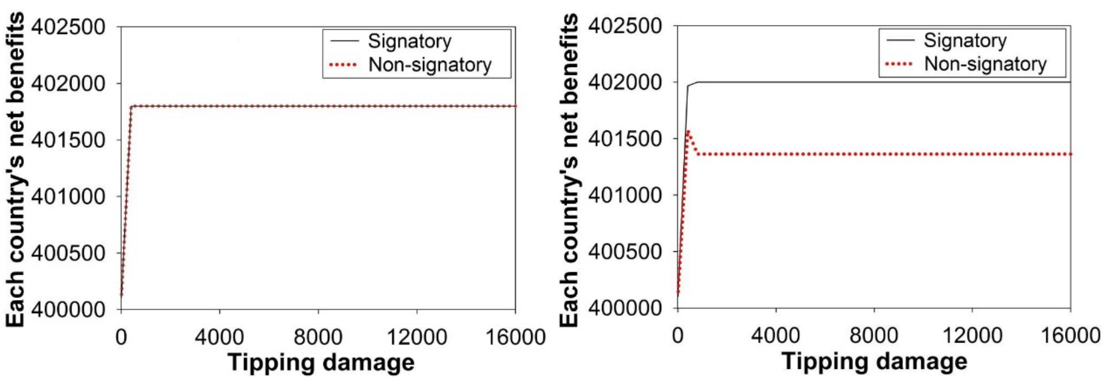
$\bar{Z}$ \ L	16000	14000	12000	10000	8000	6000	4000	2000
105000	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3
102500	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3
100000	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3
97500	None|50	None|50	None|50	None|50	None|50	None|50	None|50	None|50
95000	None|71	None|71	None|71	None|71	None|71	None|71	None|71	None|70
92500	None|87	None|87	None|87	None|87	None|87	None|87	None|87	2 or 3 or 34|2 or 3
90000	None|100	None|100	None|100	None|100	None|100	None|100	2 or 3 or 32|2 or 3	2 or 3 or 61|2 or 3
87500	None|100	None|100	None|100	None|100	None|100	2 or 3 or 39|2 or 3	2 or 3 or 60|2 or 3	2 or 3
85000	None|100	None|100	None|100	2 or 3 or 29|2 or 3	2 or 3or 48|2 or 3	2 or 3 or 64|2 or 3	2 or 3	2 or 3

Note: 'X|Y' in the table means the stable number in the Nash and Stackelberg equilibrium are 'X' and 'Y', respectively. Otherwise, the case without '|' means the stable number in the two types of equilibrium is the same. 'None' means that there is no stable coalition.

5.3. The Effect of Being the First Mover

Figure 3 shows net benefits of a signatory and a non-signatory in the two types of equilibrium. It can be seen from Figures 3(a) and (b) that, if the threshold for tipping is low but the tipping damage is not sufficient low (i.e., $\bar{Z}=98000$, $L \geq 400$), the difference of net benefits between a signatory and a non-signatory in the Stackelberg equilibrium is 638, which is much greater than that in the Nash equilibrium (where there is no difference because there is no stable coalition, i.e., all countries are non-signatories). Besides, the net benefits of a signatory in the Stackelberg equilibrium are larger than those in the Nash equilibrium, which reflects the first-mover advantage of the signatories' coalition under this circumstance.

Figure 3. Each Country's Net Benefits in the Nash and Stackelberg Equilibrium



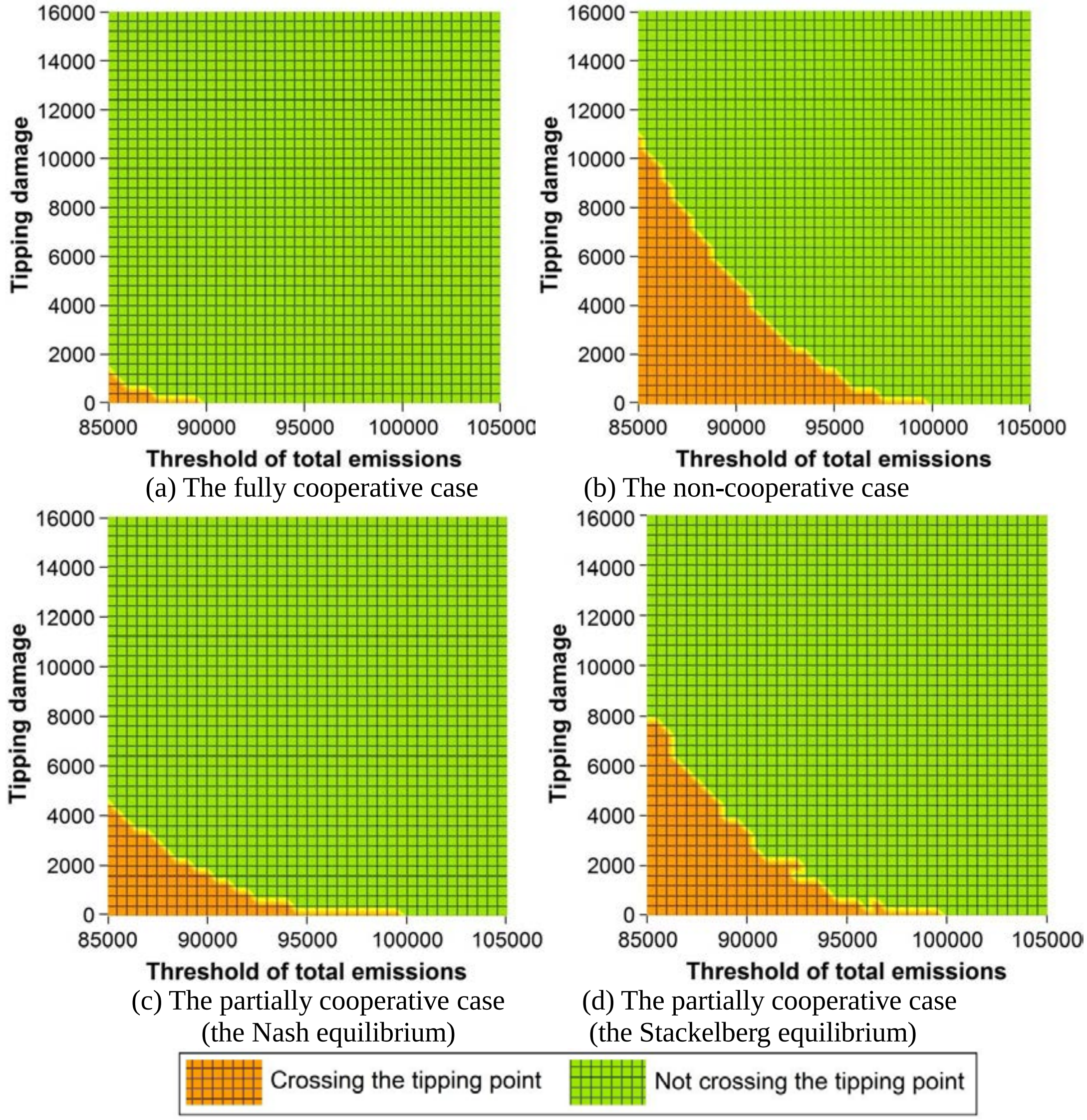
(a) The Nash equilibrium ($\bar{Z}=98000$) (b) The Stackelberg equilibrium ($\bar{Z}=98000$)

5.4. Whether or Not the Climate Tips

Fig. 4 shows whether the climate system crosses the tipping point under different circumstances. It can be seen that when, for instance, $\bar{Z}$=87500, L=6000, the climate system would cross the tipping point in the non-cooperative case, while not in the fully-cooperative or partially-cooperative cases. When $\bar{Z}$=90000, L=1000, the climate system would cross the tipping point in the partially-cooperative and non-cooperative cases, while not in the fully cooperative case. Thus, a larger scale of cooperation makes climate tipping less likely to occur, which is consistent with our prediction in the analytic results above.

Furthermore, the climate system is less likely to cross the tipping point if signatories and non-signatories make decisions simultaneously. For instance, when $\bar{Z}$=90000, L=4000, the climate system would cross the tipping point if signatories make decisions in advance, while not if they make decisions simultaneously. Thus, the signatories' coalition being the first mover may make the climate system more likely to cross the tipping point, due to the increased emissions of signatory countries compared to the Nash equilibrium.

Figure 4. Whether the Climate System Cross the Tipping Point in Various Cooperative Cases



(a) The fully cooperative case

(b) The non-cooperative case

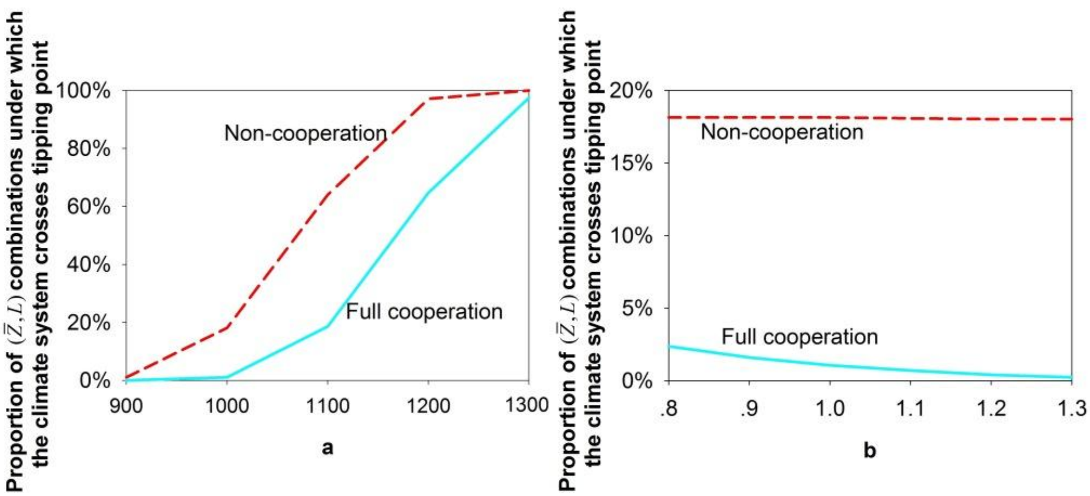
(c) The partially cooperative case (the Nash equilibrium)

(d) The partially cooperative case (the Stackelberg equilibrium)

5.5. The Effect of Benefit and Damage Parameters

Figure 5 presents the proportion of $(\bar{Z}, L)$ parameter combinations under which the climate system would cross the tipping point, with different values of 'a' and 'b' in the fully-cooperative and non-cooperative cases. It can be seen that the proportion in the non-cooperative case is always larger than that in the fully-cooperative case, no matter how 'a' and 'b' change. Thus, the values of parameters would not change the conclusion that cooperation would make climate tipping less likely to occur, which also confirms our theoretical prediction above.

Figure 5. The Effect of 'a' and 'b' on Whether the Climate System Would Cross the Tipping Point



6. Conclusions

This paper introduces the possibility of climate tipping into a standard model of IEAs to investigate its effect on GHG emissions and international cooperation. We build a game-theoretical model under various cases to analytically illustrate how the threat of climate tipping would affect individual countries' optimal emissions, their decisions about signing IEAs, and whether the climate system would cross the tipping point in equilibrium. In addition, we investigate the effect of possible tipping events on the scale of participation in IEAs on climate change in a Stackelberg equilibrium, where the signatories' coalition is the first mover.

From the theoretical model and numerical illustrations, we draw conclusions that: 1) if the threshold of total emissions for tipping is high, or the threshold is low but the tipping damage is also low, the possibility of climate tipping would have no effect on countries' emission decisions or on the size of a stable coalition; 2) if the threshold is low but tipping damage is moderate, there would be larger scale cooperation, with these circumstances promoting cooperation among countries owing to the fact that a smaller scale of cooperation might cause the climate system to cross the tipping point; 3) if the threshold is low but tipping damage is high, the threat of climate tipping may weaken cooperation among countries in the Nash equilibrium; however, the possibility of climate tipping may promote cooperation in the Stackelberg equilibrium.

In addition, we find that there might be more countries participating in cooperation when the current atmospheric carbon concentration is closer to the threshold for tipping. We also find that cooperation could decrease the possibilities of climate tipping through reducing free-riding. From these findings, we infer that a growing belief that the climate system is close to the tipping point may be one of the reasons that more countries signed the Paris Agreement than the earlier Kyoto Protocol.

However, this paper is not without limitations. For the sake of simplicity, this paper uses a static model where the countries are symmetric. In the real world, countries are different from each other. In addition, they may make decisions dynamically over time in response to the natural decay and accumulation of greenhouse gases in the atmosphere. Therefore, it is worth investigating international environmental agreements under the threat of climate tipping in a dynamic framework. In addition, relaxing the assumption of symmetric countries, which would allow the investigation of transfers among countries, can also be a direction for further research.

References

Alley, R.B., J. Marotzke, W.D. Nordhaus, J.T. Overpeck, D.M. Peteet, R.A. Pielke, Jr., R.T. Pierrehumbert, P.B. Rhines, T.F. Stocker, L.D. Talley, and J.M. Wallace. 2003. Abrupt Climate Change. *Science* 299(5615): 2005-2010.

Bahel, E. 2017. Cooperation and Subgame Perfect Equilibria in Global Pollution Problems with Critical Threshold. *Environmental and Resource Economics* (14):1-25.

Barrett, S. 1994. Self-enforcing International Environmental Agreements. *Oxford Economic Papers* 46(October Supplement): 878-894.

Barrett, S. 1995. *Heterogeneous International Environmental Agreements*. Milano: Fondazione ENI Enrico Mattei.

Barrett, S. 2013. Climate Treaties and Approaching Catastrophes. *Journal of Environmental Economics and Management* 66(2): 235-250.

Barrett, S., and A. Dannenberg. 2012. Climate Negotiations under Scientific Uncertainty. *Proceedings of the National Academy of Sciences* 109(43): 17372-17376.

Barrett, S., and A. Dannenberg. 2014. Sensitivity of Collective Action to Uncertainty about Climate Tipping Points. *Nature Climate Change* 4(1): 36-39.

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

Brozović, N., and W. Schlenker. 2011. Optimal Management of an Ecosystem with an Unknown Threshold. *Ecological Economics* 70(4): 627-640.

Cai, Y., K.L. Judd, and T.S. Lontzek. 2013. The Social Cost of Stochastic and Irreversible Climate Change (No. w18704). National Bureau of Economic Research.

Carraro, C., and D. Siniscalco. 1993. Strategies for the International Protection of the Environment. *Journal of Public Economics* 52(3): 309-328.

Chander, P., and H. Tulkens. 1995. A Core-theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution. *International Tax and Public Finance* 2(2): 279-293.

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

De Cara, S., and G. Rotillon. 2003. Multigreenhouse Gas International Agreements (No. 2003-13). THEMA (THéorie Economique, Modélisation et Applications), Université de Cergy-Pontoise.

de Zeeuw, A., and A. Zemel. 2012. Regime Shifts and Uncertainty in Pollution Control. *Journal of Economic Dynamics and Control* 36(7): 939-950.

Diamantoudi, E., and E.S. Sartzetakis. 2006. Stable International Environmental Agreements: An Analytical Approach. *Journal of Public Economic Theory* 8(2): 247-263.

Dockner, E. J., and N. V. Long. 1993. International Pollution Control: Cooperative versus Noncooperative Strategies. *Journal of Environmental Economics and Management* 25(1): 13-29.

Finus, M., and B. Rundshagen. 2001. Endogenous Coalition Formation in Global Pollution Control. *SSRN Electronic Journal* 82(1): 71-77.

Finus, M. 2002. “Game Theory and International Environmental Cooperation: Any Practical Application?” In *Controlling Global Warming: Perspectives from Economics, Game Theory and Public Choice*, edited by C. Böhringer, M. Finus, and C. Vogt. Cheltenham, UK: Edward Elgar, pp. 9-104.

Fuentes-Albero, C., and S.J. Rubio. 2010. Can International Environmental Cooperation Be Bought? *European Journal of Operational Research* 202(1): 255-264.

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

Hoel, M. 1992. International Environment Conventions: The Case of Uniform Reductions of Emissions. *Environmental and Resource Economics* 2(2): 141-159.

Iida, K. 1993. Analytic Uncertainty and International Cooperation: Theory and Application to International Economic Policy Coordination. *International Studies Quarterly* 37(4): 431-457.

IPCC. 2014a. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth*

Assessment Report of the Intergovernmental Panel on Climate Change, edited by C.B. Field, V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White. Cambridge, UK and New York, NY: Cambridge University Press, p. 1132.

IPCC. 2014b. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by R.K. Pachauri and L.A. Meyer. Geneva, Switzerland: IPCC, p. 151.

İriş, D., and A. Tavoni. 2016. Tipping Points and Loss Aversion in International Environmental Agreements. Centre for Climate Change Economics and Policy Working Paper No. 269. Grantham Research Institute on Climate Change and the Environment Working Paper No. 239.

Karp, L., and L. Simon. 2013. Participation Games and International Environmental Agreements: A Non-Parametric model. *Journal of Environmental Economics and Management* 65(2):326-344.

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

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

Lemoine, D., and C. Traeger. 2014. Watch Your Step: Optimal Policy in a Tipping Climate. *American Economic Journal: Economic Policy* 6(1): 137-166.

Lenton, T.M., H. Held, E. Kriegler, J.W. Hall, W. Lucht, S. Rahmstorf, and H.J. Schellnhuber. 2008. Tipping Elements in the Earth's Climate System. *Proceedings of the National Academy of Sciences* 105(6): 1786-1793.

Lenton, T.M. 2011. Early Warning of Climate Tipping Points. *Nature Climate Change* 1(4): 201-209.

Lessmann, K., Kornek, U., Bosetti, V. 2015. The Stability and Effectiveness of Climate Coalitions. *Environmental and Resource Economics* 62(4): 811-836.

McGinty, M. 2006. International Environmental Agreements among Asymmetric Nations. *Oxford Economic Papers* 59(1): 45-62.

McGinty, M. 2010. International Environmental Agreements as Evolutionary Games. *Environmental and Resource Economics* 45(2): 251-269.

McInerney, D., R. Lempert, and K. Keller. 2012. What are Robust Strategies in the Face of Uncertain Climate Threshold Responses? *Climatic Change* 112(3-4): 547-568.

Nkuiya, B. 2015. Transboundary Pollution Game with Potential Shift in Damages. *Journal of Environmental Economics and Management* 72:1-14.

Nkuiya, B., W. Marrouch, and E. Bahel. 2015. International Environmental Agreements under Endogenous Uncertainty. *Journal of Public Economic Theory* 17(5):752-772.

Overpeck, J.T., and J.E. Cole. 2006. Abrupt Change in Earth's Climate System. *Annual Review of Environmental Resources* 31: 1-31.

Rubio, S., and B. Casino. 2001. International Cooperation in Pollution Control. Working papers/Documentos de trabajo: Serie AD (21): 1.

Rubio, S., and A. Ulph. 2002. Leadership and Self-enforcing International Environmental Agreements with Non-negative Emissions. Discussion paper. http://www/soton.ac.uk/~econweb/dp/dp02.html

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

Russill, C., and Z. Nyssa. 2009. The Tipping Point Trend in Climate Change Communication. *Global Environmental Change* 19(3): 336-344.

Schmidt, R. C. 2017. Dynamic Cooperation with Tipping Points in the Climate System. *Oxford Economic Papers* 69(2): 388-409.

Ulph, A. 2004. Stable International Environmental Agreements with a Stock Pollutant, Uncertainty and Learning. *Journal of Risk and Uncertainty* 29(1): 53-73.

Van der Ploeg, F., and A. de Zeeuw. 2016. Non-cooperative and Cooperative Responses to Climate Catastrophes in the Global Economy: A North–South Perspective. *Environmental and Resource Economics* 65(3): 519-540.

Zhang, X.B., and L. Zhu. 2017. Strategic Carbon Taxation and Energy Pricing under the Threat of Climate Tipping Events. *Economic Modelling* 60: 352-363.

Appendices

Appendix A1

Proof of Proposition 1.

Full Cooperation

In this case, there is only one decision variable, so this is degraded to an optimization problem.

Both of the first-order conditions in two intervals are $a - nb - \frac{Q^c}{n} = 0$, so the optimal decision is $Q^c = n(a - nb)$, but it needs to satisfy some conditions.

In the interval of $Q^c \le \bar{Z}$, if $\bar{Z} \ge n(a - nb)$, the optimal decision is $Q^{c*} = n(a - nb)$; if $\bar{Z} < n(a - nb)$, the optimal decision is $Q^{c*} = \bar{Z}$. Its optimal value of net benefits is Π_1^{c*}.

In the interval of $Q^c > \bar{Z}$, if $\bar{Z} \ge n(a - nb)$, the optimal decision is $Q^{c*} = \bar{Z} + \varepsilon$; if $\bar{Z} < n(a - nb)$, the optimal decision is $Q^{c*} = n(a - nb)$. Its optimal value of net benefits is Π_2^{c*}.

Comparing Π_1^c and Π_2^c, the optimal decision is the decision which corresponds to larger benefits.

Appendix A2

Proof of Proposition 2.

Non-Cooperation.

In this case, each country needs to decide its own emissions while taking others' action as given. It is a Nash game. The first order condition is $a - q_i^{oc} - b = 0$. Using a similar approach as in Appendix A1, we can obtain the best response function of country i:

$$q_i^{oc} = \begin{cases} a - b, & Q_{-i}^{oc} \le \bar{Z} - (a - b) \\ \bar{Z} - Q_{-i}^{oc}, & \bar{Z} - (a - b) < Q_{-i}^{oc} \le \bar{Z} \,\&\, L \ge \frac{1}{2}(a - b - (\bar{Z} - Q_{-i}^{oc}))^2 \\ a - b, & \bar{Z} - (a - b) < Q_{-i}^{oc} \le \bar{Z} \,\&\, L < \frac{1}{2}(a - b - (\bar{Z} - Q_{-i}^{oc}))^2 \\ a - b, & Q_{-i}^{oc} > \bar{Z} \end{cases}$$

$$i = 1, 2, \cdots n$$

where $Q_{-i}^{oc}=\sum_{j\neq i} q_j^{oc}$.

Solving these n equations simultaneously, we can obtain the Nash equilibrium:

$$q_i^{oc*}=\begin{cases} a-b, & \bar{Z}\geq n(a-b) \\ \dfrac{\bar{Z}}{n}, & \bar{Z}<n(a-b)\ \&\ L\geq \dfrac{1}{2}(a-b-\dfrac{\bar{Z}}{n})^2 \\ a-b, & \bar{Z}<n(a-b)\ \&\ L<\dfrac{1}{2}(a-b-\dfrac{\bar{Z}}{n})^2 \end{cases}$$

Substituting the strategy into the net benefit function, we can obtain the corresponding net benefits. Now, we need to verify that this strategy is a Nash equilibrium. Assume country i changes its emissions from q_i^{oc*} to $q_i^{oc*}+\delta$, resulting in its net benefit changing from π_i^{oc} to $\pi_i^{oc\prime}$. Comparing π_i^{oc} and $\pi_i^{oc\prime}$, we find $\pi_i^{oc\prime}$ is always less than π_i^{oc}. That is, if one country changes its strategy while others remain unchanged, its net benefits would decrease. Thus, the strategy in Proposition 2 is a Nash equilibrium.

Appendix A3

Proof of Proposition 3.

Partial Cooperation (Nash Equilibrium)

In this case, we assume that signatories and non-signatories take action simultaneously. The signatories' coalition decide their members' emissions while taking non-signatories' actions as given, and each non-signatory makes decisions while taking the signatories' coalition and other non-signatories' actions as given. The first-order conditions of the coalition and a non- signatory are $a-\dfrac{Q^s}{m}-mb=0$ and $a-q_j^f-b=0$ respectively. Using a similar approach as in Appendix A2, we can obtain the best response function of the signatories' coalition and a non-signatory:

$$Q^s(m)=\begin{cases} ma-m^2b, & Q^f\leq \bar{Z}-m(a-mb) \\ \bar{Z}-Q^f, & \bar{Z}-m(a-mb)<Q^f\leq \bar{Z}\&L\geq \dfrac{1}{2}(a-mb-\dfrac{\bar{Z}-Q^f}{m})^2 \\ ma-m^2b, & \bar{Z}-m(a-mb)<Q^f\leq \bar{Z}\&L<\dfrac{1}{2}(a-mb-\dfrac{\bar{Z}-Q^f}{m})^2 \\ ma-m^2b, & Q^f>\bar{Z} \end{cases}$$

$$q_j^f(m)=\begin{cases} a-b, & Q^s+Q_{-j}^f\le \bar{Z}-(a-b) \\ \bar{Z}-Q^s-Q_{-j}^f, & \bar{Z}-(a-b)<Q^s+Q_{-j}^f\le \bar{Z}\,\&L\ge\frac{1}{2}(a-b-(\bar{Z}-Q^s-Q_{-j}^f))^2 \\ a-b, & \bar{Z}-(a-b)<Q^s+Q_{-j}^f\le \bar{Z}\,\&L<\frac{1}{2}(a-b-(\bar{Z}-Q^s-Q_{-j}^f))^2 \\ a-b, & Q^s+Q_{-j}^f>\bar{Z} \end{cases}$$

$$j=m+1,\cdots,n$$

where $Q^f=\sum_j q_j^f$ and $Q_{-j}^f=\sum_{k\neq j} q_k^f$.

Solving these $n-m+1$ equations simultaneously, we can obtain the Nash equilibria.

If $\bar{Z}<m(a-mb)+(n-m)(a-b)$, there would be many equilibria. Here, we just analyze the equilibrium in which the marginal net benefits of the signatories' coalition and a non-signatory are the same, that is, $a-\frac{Q^s}{m}-mb=a-b-q_i^f$. For instance, if tipping damage is sufficiently high, the reaction function of signatories and non-signatories is $\begin{cases} Q^s=\bar{Z}-Q^f \\ q_j^f=\bar{Z}-Q^s-Q_{-j}^f,\ j=m+1,\cdots,n \end{cases}$. Thus, $q_i^{s*}(m)=\frac{\bar{Z}+(m-n)(m-1)b}{n}$, $q_j^{f*}(m)=\frac{\bar{Z}+m(m-1)b}{n}$. Then, we can get the only Nash equilibrium with different $\bar{Z}$ and L.

We can verify this strategy is a Nash equilibrium using same method as in Appendix A2. Thus, the strategy in Proposition 3 is a Nash equilibrium.

Appendix A4

Proof of Proposition 4.

Partial cooperation (Nash Equilibrium)

With different $\bar{Z}$ and L, the equilibrium emissions of signatories are different. So, we need to verify whether the stability conditions mentioned in Section 3.3.1 (see the inequalities in (5)) hold under different $\bar{Z}$ and L, one by one. From the results in Proposition 3, we can get the value of m, which could make the stability conditions (i.e., the size of a stable coalition) hold.

It is necessary to note that, if $na-n^2b\le\bar{Z}<na-nb$, the benefits of coalitions differ with a different number of countries, resulting in different conditions for stability.

Assume that $f(m) = na - nb + mb - m^2 b$, $\frac{\partial f(m)}{\partial m} = b - 2bm < 0$ when $m \geq \frac{1}{2}$. Thus, $f(m)$ is a monotonically decreasing function of m when $m \geq \frac{1}{2}$, i.e. $f(a) < f(b)$ if $a > b \geq \frac{1}{2}$. Obviously, $f(m)$ is a continuous function, and we have $f(1) = na - nb$ and $f(n) = na - n^2 b$ Thus, $\forall \bar{Z} \in [na - n^2 b, na - nb)$, $\exists k$ ($k = 1, 2, \cdots, n-1$) which makes $\bar{Z} \in [f(k+1), f(k))$, that is, $na - nb + (k+1)b - (k+1)^2 b \leq \bar{Z} < na - nb + kb - k^2 b$. For $m = 1, 2, \cdots k$, $\bar{Z} < n(a-b) + mb - m^2 b$; while for $m = k+1, k+2, \cdots, n$, $\bar{Z} \geq n(a-b) + mb - m^2 b$.

If $\frac{1}{2}(a - nb - \frac{\bar{Z}}{n})^2 \leq L < \frac{1}{2}(a - b - \frac{\bar{Z}}{n})^2$, similarly, we know that there is always a k' ($k' = 1, 2, \cdots, n-1$) which makes L satisfy $\frac{1}{2}(a - b - \frac{\bar{Z} + k'(k'+1)b}{n})^2 \leq L < \frac{1}{2}(a - b - \frac{\bar{Z} + k'(k'-1)b}{n})^2$.

Summing up the above results, we can obtain the number of signatory countries, m, which would support a stable coalition under different $\bar{Z}$ and L in the Nash equilibrium.

Appendix A5

Proof of Proposition 5.

Partial Cooperation (Stackelberg Equilibrium)

In Stackelberg game, signatories make decisions before non-signatories. It is a two-stage game, and we use a backward induction method to solve it. First, non-signatories decide their emissions while they know the total emissions of signatories Q^s, and the $n-m$ non-signatories play a Nash game. This is similar to non-cooperative scenarios with a threshold of total emissions of $\bar{Z} - Q^s$. Thus, the equilibrium of non-signatories is

$$q_j^{f*}(m) = \begin{cases} a-b, & \bar{Z} - Q^s \geq (n-m)(a-b) \\ \frac{\bar{Z} - Q^s}{n-m}, & 0 \leq \bar{Z} - Q^s < (n-m)(a-b) \ \& \ L \geq \frac{1}{2}(a - b - \frac{\bar{Z} - Q^s}{n-m})^2 \\ a-b, & 0 \leq \bar{Z} - Q^s < (n-m)(a-b) \ \& \ L < \frac{1}{2}(a - b - \frac{\bar{Z} - Q^s}{n-m})^2 \\ a-b, & \bar{Z} - Q^s < 0 \end{cases}$$

Then, substituting the reaction function of non-signatories into the net benefit function of the signatories' coalition, and maximizing, we can obtain the optimal strategy of the signatories' coalition:

$$Q^s(m)=\begin{cases}\min(ma,\bar{Z}), & L\geq \frac{1}{2}(a-b-\frac{\bar{Z}-\min(ma,\bar{Z})}{n-m})^2\\ \bar{Z}+(n-m)[\sqrt{2L}-(a-b)], & \frac{1}{2}A(m)^2\leq L<\frac{1}{2}(a-b-\frac{\bar{Z}-\min(ma,\bar{Z})}{n-m})^2\\ m(a-mb), & L<\frac{1}{2}A(m)^2\end{cases}$$

where

$$A(m)=\frac{-(n(a-b)+mb-\bar{Z})(n-m)+\sqrt{(n(a-b)+mb-\bar{Z})^2(n-m)^2+(m^2-(n-m)^2)(n(a-b)+mb-m^2b-\bar{Z})^2}}{m^2-(n-m)^2}$$

Substituting the emissions of signatories into the reaction function of non-signatories, we can obtain the equilibrium strategy of non-signatories.

Appendix A6

Proof of Proposition 6.

Partial Cooperation (Stackelberg Equilibrium)

We adopt the same method as in Appendix A4.

Adopting the results in Proposition 5 and verifying whether the stability conditions (see the inequalities in (5)) hold under different $\bar{Z}$ and L, one by one, we can obtain the value of m that make the stability conditions hold under different $\bar{Z}$ and L, that is, the size of a stable coalition.

Summing up the above results, we can obtain the number of signatory countries, m, which would support a stable coalition in the Stackelberg equilibrium.