Int Environ Agreements (2008) 8:317–341
DOI 10.1007/s10784-008-9082-6

ORIGINAL PAPER

The formation of an international environmental agreement as a two-stage exclusive cartel formation game with transferable utilities

Porchiung Benjamin Chou · Cheickna Sylla

Accepted: 28 July 2008 / Published online: 10 September 2008
© Springer Science+Business Media B.V. 2008

Abstract We construct a two-stage exclusive cartel formation game with utility transfers to model the formation process of an international environmental agreement. Our results show that in the first stage of low degree of consensus, engaging in utility transfers by asymmetric countries will accomplish little. In contrast, in the second stage of higher degree of consensus, it is more likely for asymmetric countries to engage in monetary transfers to form the grand coalition, particularly if a small stable coalition has already been formed in the first stage. This article therefore provides a theoretical perspective to explain why it is more likely for some developed countries to initiate an IEA formation process by forming a small stable coalition first before engaging in monetary transfers to form the grand coalition with all the other countries. Such a perspective is consistent with the historical development of the Montreal Protocol and may also explain the difficulty for asymmetric countries to form the grand coalition at the beginning of the IEA formation process of the Kyoto Protocol.

Keywords Cooperative game theory · Core ·
Endogenous coalition formation with transferable utility ·
International environmental agreement · Dominant cartel formation game ·
Exclusive membership game delta

1 Introduction

For the past several decades, countries around the world have become more concerned about global environmental problems and have participated in the formation processes of international environmental agreements (IEAs), such as the Montreal Protocol and the

P. B. Chou (✉) · C. Sylla
School of Management, New Jersey Institute of Technology, University Heights,
Newark, NJ 07102, USA
e-mail: pchou@njit.edu

C. Sylla
e-mail: sylla@njit.edu

Kyoto Protocol. The 1987 Montreal Protocol and its later amendments represent a sequence of successful international cooperation and coordination, particularly with the establishment of technology and financial transfers from developed countries to developing countries. In contrast, the 1997 Kyoto Protocol was not an effective IEA, especially when the USA withdrew in 2001. Nevertheless, other countries moved on without the USA, and the Kyoto Protocol entered into force on 16 February 2005 after a sufficient number of countries signed and ratified the Kyoto Protocol. The recent development of the Bali Roadmap during the conference of the parties on its thirteenth session in December 2007 has shown even more significant and promising progress, including the establishment of the mechanisms for technology and financial transfers. In the historical development of these two IEAs, the technology and financial transfers from developed countries to developing countries have facilitated international cooperation worldwide, but at the same time have also increased the level of difficulty for all the countries to reach an agreement in a short period of time. As a result, countries usually need more than one stage to reach the consensus of all or most of the countries in an IEA formation process.

There are currently two main different perspectives of theorizing the formation of an IEA as a process of coalition formation. They are non-cooperative games with non-transferable utility, with the focus on homogeneous players, versus cooperative games with transferable utility, with the focus on (totally) heterogeneous players. However, neither perspective can fully characterize the formation process of an IEA. For example, in 1987, 24 developed countries signed the Montreal Protocol, which resembled a non-cooperative game without utility transfers. After 1990 when the amendment of the Montreal Protocol established the financial assistance and technology transfers, it became more similar to a cooperative game with utility transfers. Today, more than 190 countries have participated in the Montreal Protocol. Similarly, the Adaptation Fund under the Kyoto Protocol was not established until 2001. Additional mechanisms have also been established for the technology transfers and for the adaptation to climate changes, such as the Special Climate Change Fund and the Least Developed Countries Fund. As of April 2008, more than 175 countries have signed and ratified the Kyoto Protocol.

From a theoretical perspective of coalitional games, an important question is how and when countries should engage in utility transfers (in terms of monetary or technology transfers) over time to achieve a stable IEA and/or to expand an existing IEA. But because there are only few bridges between the cooperative and non-cooperative approaches, there is a need for more theoretical integration between these two approaches to explore the strategic interactions in the non-cooperative games where countries can engage in utility or monetary transfers. Therefore, based on a simple two-stage game theoretical model of coalition formation among two asymmetric groups of symmetric countries, namely, the developed countries and developing countries, the contribution of this article is threefold. First, our article provides an integrated approach between the cooperative and non-cooperative approaches such that countries in the non-cooperative games of endogenous coalition formation can have the option to engage in utility transfers when it is beneficial for them to do so. Second, our article builds a bridge between the approaches of homogeneous players and totally heterogeneous players. Third, because consensus building takes time, this article provides a theoretical perspective to explain why it is more likely for some developed countries to initiate an IEA formation process by forming a small stable coalition first before engaging in monetary transfers to form the grand coalition with all the other countries.

Furthermore, for the methodology of non-cooperative game theory, countries engage in endogenous coalition formation under different rules of the game, in both simultaneous games and sequential games.[1] Many authors have contributed to this approach, such as Carraro and Siniscalco (1991, 1993), Carraro et al. (2006), Barrett (1994, 1997), Bloch (1995, 1996, 1997), Yi (1997), Finus (2003b), Finus and Rundshagen (2003b, c), Eyckmans and Finus (2006), and Caparrós et al. (2004). However, with the exception of Finus and Rundshagen (2003b), Carraro and Siniscalco (1993), and Carraro et al. (2006), in the non-cooperative approach, there are no utility transfers within or across coalitions. In addition, as the results that can be derived analytically are usually based on the framework of n symmetric countries, including the article of Carraro and Siniscalco (1993), an important research question is whether their key results remain robust for asymmetric or heterogeneous countries, particularly when countries can engage in monetary transfers.

For the methodology of cooperative games, countries participate in forming the grand coalition via a sharing rule that specifies utility or monetary transfers among countries, proposed by Chander and Tulkens (1992, 1995, 1997). They also propose the γ-core characteristic function as the concept underlying the sharing rule, which implies that if a country or some countries deviate from the grand coalition, the rest of the coalition breaks up into singletons to maximize their own individual payoffs. The sharing rule proposed by Chander and Tulkens (1992, 1995, 1997) has also been applied and extended by many researchers, such as Botteon and Carraro (1997), Tulkens (1998), Chander et al. (1999), Germain et al. (1997, 2003), Germain and van Ypersele (1999), Chander and Tulkens (2001, 2006), Chander (2007), Eyckmans and Tulkens (2003), Eyckmans and Finus (2004), Figuières and Magali (2003), Weikard (2005), Weikard et al. (2006), Chou and Shogren (2008), and Brèchet et al. (2007). However, as Tulkens (1998) points out, the theory of the (γ-core) cooperative game is not the theory of endogenous coalition formation and does not characterize *how* countries actually form the grand coalition. In addition, when countries are totally heterogeneous, it is difficult to derive results analytically without resorting to simulations, or ad hoc assumptions about how countries form coalitions.

Several authors, such as Tulkens (1998), Finus and Rundshagen (2003a, b), Carraro et al. (2006), Chander (2007), Chander and Tulkens (2006), Weikard (2005), Brèchet et al. (2007), and Chou and Shogren (2008), have contributed to the few bridges between these two approaches. While some of them, such as Tulkens (1998) and Chander and Tulkens (2006), continue to treat the cooperative and non-cooperative approaches as two separate methodologies to compare and contrast; others, such as Chander (2007) and Finus and Rundshagen (2003a), focus on exploring the non-cooperative foundations of the core concept. At the same time, there are also authors who focus on applying the sharing rule to the non-cooperative games, such as Finus and Rundshagen (2003b), Carraro et al. (2006), Weikard (2005), and Chou and Shogren (2008).[2] While these authors focus on the applications to heterogeneous countries, our focus is on the applications of the sharing rule to asymmetric countries, which allows us to derive results analytically, and thus strengthen the bridge between the frameworks of homogeneous countries and totally heterogeneous countries. This is one of the most important contributions of our article.

[1] The main simultaneous games are the dominant cartel formation game (DCFG), (restricted) open membership game, exclusive membership game Δ, and exclusive membership game Γ. The main sequential games are sequential move unanimity game, equilibrium binding agreement game, and DCFG under equilibrium binding agreement. See, for example, Finus and Rundshagen (2003b, c) for details. Also see Finus (2003a) for a survey of different non-cooperative and cooperative approaches.

[2] The rules of non-cooperative games with utility transfers are similar to those in Finus and Rundshagen (2003b), although they did not discuss the dominant cartel formation game.

From the perspective of the non-cooperative games, one challenging task is to maintain the stability of a larger coalition or the grand coalition. While some authors, such as Carraro et al. (2006), use different kinds of compensation schemes, others resort to different rules of the game with a stronger punishment for countries that deviate from the large or grand coalition. Our approach focuses on different rules of the game, but at the same time includes different possible ways of monetary transfers. In particular, we construct a game that combines the dominant cartel formation game, originally proposed by d'Aspremont et al. (1983), and the exclusive membership game Δ, proposed by Hart and Kurz (1983). We refer to it as the exclusive cartel formation game (ECFG). In such an ECFG, some countries participate in forming only one non-trivial or non-singleton coalition, but with the option to exclude a new member. Each non-participating country forms only a trivial or singleton coalition. More importantly, when countries jointly form the only non-trivial coalition in the ECFG, they always have the option to engage in utility transfers.

This article includes five additional sections. Section 2 provides the preliminaries of the model, as well as a sharing rule for monetary transfers. Section 3 discusses the case in which countries engage in utility transfers in the first stage of an IEA formation process, when consensus is needed only among a small number of participating countries of an IEA. Section 3 also shows that engaging in monetary transfers in the first stage will accomplish little and that it is more efficient for three developed countries to form a small stable coalition with zero or no monetary transfers. Section 4 discusses the case in which all countries form the grand coalition in the second stage of the IEA formation process, after three developed countries have already formed a small stable coalition in the first stage. Because the agreements of all the countries are required to form the grand coalition, our results show that the developed countries, in particular those inside the only non-trivial coalition, have to offer side payments to developing countries such that the small stable coalition can be expanded to the grand coalition. Important policy implications are also provided in Sect. 4 regarding *how* and *when* countries should engage in utility transfers in the process of the IEA formation. We further discuss the strategy combination of all the countries that constitutes a subgame perfect Nash equilibrium. Section 5 discusses some possible extensions for future research. Section 6 provides summary and concluding remarks. The detailed proofs of all lemmas and propositions are provided in Appendix.

2 Preliminaries

2.1 Players, coalitions, and utility functions

Assume that there are only two types of countries in the world. Type-I countries have high GDP or GDP per capita, high emissions of pollution, stronger preferences for environmental quality than for economic development, and high abatement costs. The type-J countries have low GDP or GDP per capita, low emissions of pollution, weaker preferences for environmental quality than for economic development, and low abatement costs. Although there are exceptions, the type-I countries are usually the industrially advanced countries or developed countries while the type-J countries are usually the less-developed countries or the developing countries, which are our focus.[3]

The player set for all the developed countries is $M = \{i | i = 1,2,\ldots,m\}$, and the player set for all the developing countries is $N = \{j | j = 1,2,\ldots,n\}$, with $n > m > 3$. A coalition is

[3] See Shogren (1999) or Brèchet et al. (2007) for similar perspectives.

a non-empty subset of the player set, $M \cup N$. A singleton or trivial coalition consists of only one player in the coalition. When there is more than one player in a coalition, it is called a non-singleton or non-trivial coalition. The grand coalition is a coalition formed by all the players in the player set. In addition, a coalition structure is a partition of the player set such that the intersection of any two non-empty subsets of the player set is the empty set, Ø; the unions of all the subsets of the player set constitute the player set.

Assume that there are p out of m developed countries and q out of n developing countries that jointly form a non-trivial coalition if it is beneficial for them to do so, with $0 \leq p \leq m$, $0 \leq q \leq n$, $p + q \geq 2$. The numbers of participating p developed countries and q developing countries are *endogenous* and depend on the rules of the game. The player set for the developed countries that form the non-trivial coalition is $P = \emptyset \cup \{i | i = 1,2,\ldots,p\}$, and the player set for the developing countries that join the non-trivial coalition is $Q = \emptyset \cup \{j | j = 1,2,\ldots,q\}$, with $P \subseteq M$ and $Q \subseteq N$.

In addition, because the developed countries and developing countries are assumed to be symmetric within their own groups or types of countries, in this article, we use the symbol (p,q) to denote the p developed countries and q developing countries that jointly form the only non-trivial coalition, $\{(p,q)\}$, while all the other $(m - p)$ developed countries and $(n - q)$ developing countries form only singletons.

The net benefit or utility function for a developed country is $U_i = B_i\ (X) - C_i\ (x_i) = \alpha X - (1/2)x_i^2$, with $\alpha > 1$. For a developing country, $U_j = B_j\ (X) - C_j\ (x_j) = X - (1/2)x_j^2$, where X is the aggregate level of public good. Hence, the developed countries always value the public good more than do the developing countries. As for the cost function, for simplicity, we assume that the cost functions are identical between the developed and developing countries. However, just because countries have the same cost functions, it does not imply that they have the same abatement costs, which depend on the levels of abatements.

In addition, as Bloch (1997) points out, it is a natural assumption that players inside a non-trivial coalition cooperatively maximize the aggregate payoff of the coalition, and among non-trivial coalitions and singleton coalitions, players behave non-cooperatively. The goal of the p developed and q developing countries inside a non-trivial coalition is to jointly maximize the aggregate level of payoffs of the coalition, denoted by the objective function $\left[\sum_{i \in P} U_{i,c} + \sum_{j \in Q} U_{j,c}\right]$. The public goods provided by each of these p developed countries and q developing countries are denoted by $x_{i,c}$ and $x_{j,c}$. If countries form singletons or trivial coalitions, they maximize their own individual payoffs, denoted by $U_{i,n}$ and $U_{j,n}$, for the developed and developing countries, respectively. The public goods provided by the non-participating countries of the IEA are denoted by $x_{i,n}$ and $x_{j,n}$, respectively, for each of the $(m - p)$ developed countries and $(n - q)$ developing countries.

Note that the specifications of utility functions imply the orthogonal reaction functions, which, by tradition in the literature, can be interpreted as the dominant strategy. That is, when countries maximize their utilities, either in a non-trivial coalition or in a singleton coalition, the optimal levels of public good that they should provide do not depend on other countries' actions or reactions. Therefore, when some countries form a non-trivial coalition to maximize their aggregate utilities, they can anticipate that other countries will continue to form singletons to maximize their own individual utilities.[4]

[4] Most economists today believe it is not inappropriate to use the linear benefit or damage functions. See Chander and Tulkens (1995) and Na and Shin (1998) for further discussions.

2.2 A sharing rule of aggregate net benefits

In this section, we provide a sharing rule for countries to engage in utility or monetary transfers in the IEA formation process. Before countries engage in utility transfers, let the utility of a developed country be denoted by V_i, and the utility of a developing country be denoted by V_j. These utility levels are usually, but not limited to, the utility levels of players under a Nash equilibrium. More importantly, these utilities, V_i and V_j, are *not* limited to the situation when all countries form singletons, as in Chander and Tulkens (1995). In a similar manner, let the utility of a participating developed country be W_i, and the utility of a participating developing country be W_j, after countries engage in the sharing rule to form a larger coalition. But countries do not always have to form the grand coalition after they engage in utility transfers. Such a sharing rule to determine the monetary transfers among the participating countries offers more flexibility than the sharing rule proposed by Chander and Tulkens (1995) and is similar to those proposed by Finus and Rundshagen (2003c) and applied by Brèchet et al. (2007).

For the participating countries of the non-trivial coalition $\{(p,q)\}$, they have the option to engage in monetary or utility transfers via the sharing rule. Most importantly, both p and q are endogenous variables, and so are the monetary transfers among these participating countries. All these endogenous variables have to be determined at the same time. The participating countries that jointly form a non-trivial coalition $\{(p,q)\}$ have the following sharing rule for monetary transfers.

Definition 2.1 $\tau_i = -[W_i - V_i] + [\alpha/(\alpha p + q)][\Delta\text{ANB}]$, for a developed country, and $\tau_j = -[W_j - V_j] + [1/(\alpha p + q)][\Delta\text{ANB}]$, for a developing country.

The change or increase in aggregate net benefit is $\Delta\text{ANB} = [W - (\sum_{i \in P} V_i + \sum_{j \in Q} V_j)] > 0$, where $W = \sum_{i \in P} W_i + \sum_{j \in Q} W_j$.

That is, in the calculations of the monetary transfers, all the participating countries contribute their individual gains or net benefits (in monetary terms) obtained from forming the non-trivial coalition $\{(p,q)\}$. Each developed country contributes $[W_i - V_i]$, while each developing country contributes $[W_j - V_j]$. Then, all the participating countries share the increase in the aggregate gains or net benefits, $\Delta\text{ANB} = [W - (\sum_{i \in P} V_i + \sum_{j \in Q} V_j)] = [\sum_{i \in P} (W_i - V_i) + \sum_{j \in Q} (W_j - V_j)]$, according to the proportion of their individual marginal benefits out of the aggregate marginal benefits of all the participating countries. One condition for the countries to engage in utility transfers is that $\Delta\text{ANB} > 0$ must hold. Therefore, given the numbers of p developed countries and q developing countries, each developed country has a share $[\alpha/(\alpha p + q)]$ out of $[W - (\sum_{i \in P} V_i + \sum_{j \in Q} V_j)]$, while each developing country has a share $[1/(\alpha p + q)]$ out of $[W - (\sum_{i \in P} V_i + \sum_{j \in Q} V_j)]$. When a country gains more than it contributes, it is a net receiver of monetary transfers. When a country contributes more than it gains, it is a net contributor of monetary transfers. At the end, the summation of all the monetary transfers must equal zero. That is, $\sum_{i \in P} \tau_i + \sum_{j \in Q} \tau_j = 0$.

In addition, since the benefits can be canceled out in the formula, there is an alternative and sometimes more convenient way of expressing the sharing rule, as shown by Chander and Tulkens (1995). For the p developed countries, let $C_i(V)$ and $C_i(W)$ represent the abatement costs before and after these countries form the non-trivial coalition by engaging in utility transfers. For the q developing countries, let $C_j(V)$ and $C_j(W)$ represent the abatement costs before and after countries form the non-trivial coalition by engaging in utility transfers. Then for the p developed countries and q

developing countries, we have the following alternative sharing rule for monetary transfers.

Definition 2.2 $\tau_i = -[C_i(V) - C_i(W)] + [\alpha/(\alpha p + q)][\Delta\mathrm{AC}]$, for a developed country, and $\tau_j = -[C_j(V) - C_j(W)] + [1/(\alpha p + q)][\Delta\mathrm{AC}]$, for a developing country, where the change in aggregate cost = $\Delta\mathrm{AC} = [\sum_{i\in P} C_i(V) + \sum_{i\in Q} C_j(V)] - [\sum_{i\in P} C_i(W) + \sum_{j\in Q} C_j(W)]$.

In this expression, the interpretation is slightly different. Each country is compensated by the increased amount of abatement cost, while at the same time each country has to share the burden of aggregate increased abatement costs of all the countries. Because $\alpha > 1$, a developed country benefits more from the formation of the non-trivial coalition $\{(p,q)\}$ than a developing country. But at the same time, a developed country would have to share more of the increased aggregate abatement costs than would a developing country.

However, if all the participating countries that engage in the sharing rule are *ex ante* symmetric, which means that these symmetric countries all form the same coalition or coalitions, including, but not limited to, singletons, the sharing rule will generate only zero monetary transfers.[5] Therefore, in this article, when *ex ante* symmetric countries form a non-trivial coalition, it can be perceived either as a case of no utility transfers or as a special case of zero utility transfers.

2.3 Rules of the non-cooperative games and the stability conditions

There are two stages in the ECFG as an IEA formation process. In both stages, countries always have the option to engage in utility transfers. The first stage represents the early stage of an IEA formation process when there is a low degree of consensus among countries, while the second stage represents the later stage of the same IEA formation process when there is a high degree of agreement among countries. In both stages, countries form only *one* non-trivial coalition that represents the IEA, regardless of whether countries engage in utility transfers. Therefore, each country has two choices, to form a singleton or to jointly form a non-singleton coalition with other countries. But a non-trivial coalition can be formed only if all the participating countries can reach an agreement on which countries should jointly form the non-trivial coalition.

Moreover, at the end of both stages, the non-trivial coalitions formed must be stable or at least internally stable. Internal stability means that no country inside a non-singleton coalition has the incentive to deviate. External stability means that no country outside a non-singleton coalition has the incentive to join. And a coalition is stable only if it is both internally and externally stable. The satisfaction of both the internal and external stability conditions implies that the strategy combination of all the players is a Nash equilibrium, while the internal or external stability alone is not. The only exception is the grand coalition, which requires only the internal stability condition.

Before the first stage begins, all countries form singletons and maximize their own individual utilities. Although the strategy combination is a Nash equilibrium, it represents the situation of no IEA. Then, the first stage of an IEA formation starts. At the beginning of the first stage, all countries choose a message from its message space that contains all the possible coalitions that a county can form a coalition with. A country includes only the countries in the message that it wants to form a coalition with, and at the same time excludes all the other countries that it does not intend to form a coalition with. If a country intends to (continuously)

[5] See Lemma 2.1 in Appendix for details.

form a singleton in the first stage, it announces a message that contains only the country itself.[6] In addition, only the countries that announce the same message can form the only non-trivial coalition. If no two countries announce the same message, then the process of announcing messages continues until a stable or internally stable coalition is formed. If more than one non-trivial coalition is formed at the same time, then countries readjust their messages until one non-singleton coalition is left since countries can always choose to merge coalitions or to form singletons. The first stage of the game ends, when a stable non-singleton coalition is established. Such a rule of the game implies that if a country deviates from a non-trivial coalition, the rest of the coalition does *not* dissolve into singletons. Because of this, countries have the incentive to be free riders by forming singletons to maximize their own individual utilities. Therefore, the non-trivial coalition formed in the first stage usually has only a small number of participants, even if they can engage in monetary transfers.

Take a four-player game as an example. The message space for player 1, for example, is $M_1 = \{\{1\},\{1,2\},\{1,3\},\{1,4\},\{1,2,3\},\{1,2,4,\},\{1,3,4\},\{1,2,3,4\}\}$. If the messages announced by the players are $m_1 = \{1,2,3\}$, $m_2 = \{1,2,3\}$, $m_3 = \{1,2,3,4\}$, and $m_4 = \{4\}$, this means that both players 1 and 2 want to form a coalition $\{1,2,3\}$ with player 3, but at the same time exclude player 4. Player 3 wants to form the grand coalition $\{1,2,3,4\}$ with all the other players. But player 4 wants to form only a singleton and thus exclude all the other players. In this case, only players 1 and 2 can form the non-trivial coalition $\{1,2\}$, while players 3 and 4 form singletons, $\{3\}$ and $\{4\}$. Also, if players 1 and 2 prefer to form a coalition with just each other, they can simply announce the message $m_1 = m_2 = \{1,2\}$, thus excluding players 3 and 4. To relate this simple example to the context of this article, we can assume that players 1 and 2 are the European (or Scandinavian) countries, player 3 is the USA, and player 4 is a developing country. As in the dilemma of the Kyoto Protocol, USA would not participate in the treaty unless the developing country is involved. However, even without the participation of the USA, European countries can still reach an agreement and form a small coalition in the first stage of an IEA formation process. Alternatively, one may also argue that for the European countries, $m_1 = m_2 = \{1,2,3,4\}$. For the USA, $m_3 = \{3\}$, and for the developing country, $m_4 = \{4\}$. Then, European countries can still form the small coalition without the USA in the first stage.

Given that a stable non-trivial coalition is achieved in the first stage, the second stage starts when countries start to expand this non-trivial coalition to a larger coalition. That is, at the beginning of the second stage, countries start re-announcing their messages to expand the small coalition to a larger coalition or the grand coalition. The second stage ends when a stable grand coalition is achieved. In the coalition expansion process, countries always have the option to engage in utility transfers. It is possible that the second stage can be a long and gradual process. For reasons of simplicity, in Sect. 4, we focus on only one possibility when countries engage in utility transfers to expand the small coalition directly to the grand coalition. Consistent with the argument of Carraro and Sinisclaco (1993),[7] such a coalition expansion process with utility transfers must be a Pareto-improvement for all the countries, which means no countries are worse off and at least one country is strictly better off. By using the same example of a four-player game, the grand coalition can be formed only if

[6] The original dominant cartel formation game, proposed by d'Aspremont et al. (1983), becomes a special case of the ECFG if all the countries that intend to form the non-trivial coalition send the same message that includes all the countries in the player set, while countries that intend to form singletons send the message that contains only their own individual countries.

[7] Finus (2003a) and Carraro et al. (2006) refer to such Pareto-improving financial transfers as *ex post* transfers.

$m_1 = m_2 = m_3 = m_4 = \{1,2,3,4\}$. At the same time, if countries engage in monetary transfers to form a (larger) coalition, all the participating countries must agree on the amounts of monetary transfers.

When all countries form the grand coalition, its (internal) stability can be maintained only if there is a strong punishment, based on the rules of the game, for the deviating country or countries. One possibility is that if there is any deviation from the grand coalition, the countries that have originally formed the small stable coalition will re-announce the messages and resume the original small stable coalition, which is our focus in Sect. 4. There are other rules of the game that imply even more severe punishment, such as exclusive membership game Γ, originally proposed by von Neumann and Morgenstern (1944) and reintroduced by Hart and Kurz (1983), or the γ-core concept, proposed by Chander and Tulkens (1995). Under the rules of these two games, if a country or countries deviate from the grand coalition, the rest of the coalition breaks up into singletons as a punishment to the deviating country or countries. Because countries do not gain by deviating from the grand coalition, the grand coalition is (internally) stable.

In this article, when the grand coalition is achieved, the strategy combination of all the countries constitutes a Nash equilibrium in the second stage. This is simply because if any country deviates from the current strategy combination that sustains the grand coalition with monetary transfers, then the deviating country, along with all the other countries, will be worse off because the alternative is the small stable coalition. In Sect. 4.2, we will further discuss the strategy combination of countries in both stages that constitutes a subgame perfect Nash equilibrium.

2.4 Time consistency and perfectly observable actions

By following the traditional assumption made by Chander and Tulkens (1995) and all the other authors that apply the sharing rule, a participating developing country with utility transfers will lose the side payment if it deviates from the non-trivial coalition formed via the sharing rule of utility transfers, which is equivalent to the situation of not participating in the coalition in the first place. Thus, a developing country cannot receive the side payment first and then deviate. Similarly, for the participating developed countries, they cannot join a non-trivial coalition and refuse to pay for the monetary transfers. Therefore, all the countries must be time-consistent, which rules out the problem of time inconsistency in this article. From a policy perspective, time consistency to some extent can be sustained if the side payments are transferred gradually, for example, on a per-project basis, instead of a lump-sum payment. Another important assumption is that the actions of the countries in a non-trivial coalition are perfectly observable to all the other participating countries such that no countries can sign the treaty and then cheat by maximizing their individual utilities in their private actions.

3 The first stage of an IEA formation process

3.1 ECFG with utility transfers

The goal of countries in the first stage is to form a stable non-singleton coalition from the situation when all countries form singletons. Since all countries form singletons before the beginning of the first stage, $V_i = U_{i,n}$ (0,0) and $V_j = U_{j,n}$ (0,0)[8] are the utilities of a

[8] Note that $U_{i,n}$ (0,0) = $U_{i,n}$ (1,0) = $U_{i,n}$ (0,1) = $U_{i,c}$ (1,0) and $U_{j,n}$ (0,0) = $U_{j,n}$ (1,0) = $U_{j,n}$ (0,1) = $U_{j,c}$ (0,1).

developed country and a developing country, respectively, when all countries form singletons. Also, $W_i = U_{i,c}$ (p,q) and $W_j = U_{j,c}$ (p,q) are the utilities of a developed country and a developing country that jointly form the only non-singleton coalition, although both p and q are endogenous and need to be determined.

From Definition 2.1, the monetary transfers are defined as follows:

Definition 3.1 In the first stage, the monetary transfers to form a non-trivial coalition $\{(p,q)\}$, with $0 \leq p \leq m$, $0 \leq q \leq n$, and $p + q \geq 2$, are $\tau_i = -[U_{i,c}(p,q) - U_{i,n}(0,0)] + [\alpha/(\alpha p + q)][W - (\sum_{i\in P} U_{i,n}(0,0) + \sum_{j\in Q} U_{j,n}(0,0))]$, for a participating developed country, and $\tau_j = -[U_{j,c}(p,q) - U_{j,n}(0,0)] + [1/(\alpha p + q)][W - (\sum_{i\in P} U_{i,n}(0,0) + \sum_{j\in Q} U_{j,n}(0,0))]$, for a participating developing country.

We can also identify the net contributors and net receivers of monetary transfers in the following proposition for the participating countries.

Proposition 3.1 *For any combinations of p developed countries and q developing countries that jointly form a non-trivial coalition* $\{(p,q)\}$, $1 \leq p \leq m$, $1 \leq q \leq n$, *under the sharing rule, the developed countries are always the net contributors of the monetary transfers, while the developing countries are always the net receivers.*

That is, under such a sharing rule, the developed countries are always the net contributors of monetary transfers ($\tau_i < 0$), while the developing countries are always the net receivers ($\tau_j > 0$), for any combination of p and q, $1 \leq p \leq m$, $1 \leq q \leq n$. Hence, the underlying foundation for countries to employ the sharing rule to form $\{(p,q)\}$ is that the participating developed countries agree to offer side payments to the participating developing countries to form a non-trivial coalition. At the same time, these side payments must sufficiently compensate the developing countries such that they agree to join the non-trivial coalition.

Recall that for the participating countries, the payoffs with monetary transfers are $Y_i(p,q) = U_{i,c}(p,q) + \tau_i$ for a developed country and $Y_j(p,q) = U_{j,c}(p,q) + \tau_j$ for a developing country. In Lemma 3.2 (in Appendix), we show that both Y_i and Y_j are increasing functions of p and q. That is, $\frac{\partial Y_i(p,q)}{\partial p} > 0$, $\frac{\partial Y_i(p,q)}{\partial q} > 0$, $\frac{\partial Y_j(p,q)}{\partial p} > 0$, and $\frac{\partial Y_j(p,q)}{\partial q} > 0$. This means that the greater the number of participants, the higher the utility levels for the participating countries. Hence, the non-trivial coalition should be as large as possible if the participating countries want to maximize their own utilities. However, the problem of free riding arises because non-participating countries benefit even more from the formation of the non-trivial coalition. The internal stability and external stability conditions therefore determine the numbers of developed and developing countries forming the non-trivial coalition in the first stage.

The internal stability condition means that none of the participating countries of the non-trivial coalition has an incentive to deviate. Mathematically, this means that $Y_i(p,q) = U_{i,c}(p,q) + \tau_i \geq U_{i,n}(p-1,q)$, for a participating developed country, and $Y_j(p,q) = U_{j,c}(p,q) + \tau_j \geq U_{j,n}(p,q-1)$, for a participating developing country.[9] We characterize an internally stable non-trivial coalition in the following propositions.

Proposition 3.2 *If a developing country in a non-trivial coalition does not have the incentive to deviate, neither does a developed country in the same non-trivial coalition.*

[9] If the equality holds, countries are assumed to join the non-trivial coalition rather than form singleton coalitions.

That is, the satisfaction of the internal stability of a developing country inside a non-trivial coalition implies the satisfaction of the internal stability of a developed country inside the same coalition. This is simply because developed countries benefit much more from jointly forming a non-trivial coalition with the developing country or countries. Therefore, a non-trivial coalition jointly formed by both types of countries is internally stable if the developing country or countries inside the non-trivial coalition do not have the incentive to deviate. In addition, when $(p,q) \geq (1,1)$, the following proposition holds:

Proposition 3.3 *While both coalitions* $\{(p,q)\} = \{(1,1)\}$ *and* $\{(p,q)\} = \{(2,1)\}$ *are internally stable for the developed countries, only the coalition* $\{(p,q)\} = \{(1,1)\}$ *is internally stable for the developing country.*

In addition, the external stability condition means that none of the non-participating countries has an incentive to join the non-trivial coalition. Mathematically, this means that $Y_i\ (p+1,q) = U_{i,c}\ (p+1,q) + \tau_i < U_{i,n}\ (p,q)$ for a non-participating developed country, and $Y_j\ (p,q+1) = U_{j,c}\ (p,q+1) + \tau_j < U_{j,n}\ (p,q)$ for a non-participating developing country. Hence, we have the following proposition for the non-participating developed and developing countries.

Proposition 3.4 *While both coalitions* $\{(p,q)\} = \{(1,1)\}$ *and* $\{(p,q)\} = \{(2,1)\}$ *are externally stable for the developing countries, only the coalition* $\{(p,q)\} = \{(2,1)\}$ *is externally stable for the developed countries.*

Therefore, from the perspective of the developing countries, the coalition $\{(p,q)\} = \{(1,1)\}$ is both internally and externally stable. From the perspective of the developed countries, if $\{(p,q)\} = \{(1,1)\}$ is formed, one developed country should have the incentive to join the non-trivial coalition such that $\{(p,q)\}$ becomes $\{(2,1)\}$. However, if $\{(p,q)\} = \{(2,1)\}$ occurs, the developing country in the coalition will immediately deviate such that $\{(p,q)\} = \{(2,0)\}$. Because of this, the developed countries have two choices. The first choice is to let a developed country and a developing country form the non-trivial coalition with monetary transfers. This means that $\{(p,q)\} = \{(1,1)\}$ since it is both internally and externally stable for the developing countries. The second choice for the developed countries is to form the non-trivial coalition without a developing country. From the equilibrium discussed by Chou and Sylla (2006), the small coalition $\{(p,q)\} = \{(3,0)\}$ is both internally and externally stable for the developed country. It is also externally stable for the developing countries, since no developing country has the incentive to participate in this coalition. Between the $\{(p,q)\} = \{(3,0)\}$ and $\{(p,q)\} = \{(1,1)\}$, the following proposition illustrates which non-trivial coalition is more efficient.

Proposition 3.5 *It is more efficient for three developed countries to form the non-trivial coalition among themselves, i.e.,* $\{(p,q)\} = \{(3,0)\}$, *in the first stage of the ECFG.*

That is, because $U_{i,c}\ (3,0) > Y_i\ (1,1)$, $U_{j,n}\ (3,0) > Y_j\ (1,1)$, $U_{i,n}\ (3,0) > U_{i,c}\ (3,0) > Y_i\ (1,1)$, *all* the developed and developing countries are better off when three developed countries form the non-trivial coalition with zero or no transferable utility, rather than the situation where one developed country and one developing country jointly form the only non-trivial coalition with non-zero transferable utility. Hence, developed countries can accomplish very little if one developed country engages in non-zero utility transfers with one developing country to form a non-trivial coalition in the first stage of an IEA formation process. Therefore, developed countries should choose to form the only non-trivial coalition $\{(3,0)\}$ among themselves with zero or no monetary transfers to initiate the IEA, instead of trying to form a non-trivial coalition with a developing country with utility transfers. Thus, only {(3,0)} will be formed, instead of {(1,1)}, and there is no movement from {(1,1)} to {(3,0)}. In addition, this means that three developed countries only need to announce a message to include one another to form the stable non-trivial coalition. At the same time, all the other developed countries and developing countries should announce the messages that include only themselves, since they can be better off by forming singletons to free ride the efforts of the three developed countries.

4 The second stage of an IEA formation process

In the second stage, the goal of countries is to expand the small coalition to the grand coalition. Given that some countries have already formed a small non-trivial coalition $\{(p,q)\}$ in the first stage, all the countries can now engage in utility transfers to expand the small stable coalition to the grand coalition.

4.1 The sharing rule with a pre-existing small stable coalition

Let us continue to assume that the existing stable small coalition consists of $\{(p,q)\}$, with $0 \leq p \leq m$, $0 \leq q \leq n$, and $2 \leq p + q < m + n$. For the countries originally in the small coalition, $\{(p,q)\}$, $V_i = U_{i,c}\ (p,q)$, $V_j = U_{j,c}\ (p,q)$, $W_i = U_{i,c}\ (m,n)$, and $W_j = U_{j,c}\ (m,n)$. Also, for the countries that originally stay outside of the small coalition, $V_i = U_{i,n}\ (p,q)$, $V_j = U_{j,n}\ (p,q)$, $W_i = U_{i,c}\ (m,n)$, and $W_j = U_{j,c}\ (m,n)$. The monetary transfers that countries can use to expand the small coalition to the grand coalition are in the following definition.

Definition 4.1 For the countries originally in the small coalition, $\tau_{i,c} = -[U_{i,c}(m,n) - U_{i,c}(p,q)] + [\alpha/(\alpha m + n)][\Delta \text{ANB}]$, for a developed country, and $\tau_{j,c} = -[U_{j,c}\ (m,n) - U_{j,c}(p,q)] + [1/(\alpha m + n)][\Delta \text{ANB}]$, for a developing country.

For the countries that are originally outside the small coalition, $\tau_{i,n} = -[U_{i,c}\ (m,n) - U_{i,n}\ (p,q)] + [\alpha/(\alpha m + n)][\Delta \text{ANB}]$, for a developed country, and $\tau_{j,n} = -[U_{j,c}\ (m,n) - U_{j,n}\ (p,q)] + [1/(\alpha m + n)][\Delta \text{ANB}]$, for a developing country.

The change in aggregate net benefit = $\Delta \text{ANB} = \left[W - \left(\sum_{i \in P} U_{i,c}(p,q) + \sum_{j \in Q} U_{j,c}(p,q) + \sum_{i \in M \backslash P} U_{i,n}(p,q) + \sum_{j \in N \backslash Q} U_{j,n}(p,q)\right)\right]$, with $\Delta \text{ANB} > 0$.[10]

Similar to Definition 2.2, we have the following alternative definition.

Definition 4.2 For the countries that originally participate in the small coalition, $\tau_{i,c} = [C_i\ (x_{i,c}\ (m,n)) - C_i\ (x_{i,c}\ (p,q))] + [\alpha/(\alpha m + n)][\Delta \text{AC}]$, for a developed country, and $\tau_{j,c} = [C_j\ (x_{j,c}\ (m,n)) - C_j\ (x_{j,c}\ (p,q))] + [1/(\alpha m + n)][\Delta \text{AC}]$, for a developing country.

For the countries that do not participate in the small coalition, $\tau_{i,n} = [C_i\ (x_{i,c}\ (m,n)) - C_i\ (x_{i,n}\ (p,q))] + [\alpha/(\alpha m + n)][\Delta \text{AC}]$, for a developed country, and $\tau_{j,n} = [C_j\ (x_{j,c}\ (m,n)) - C_j\ (x_{j,n}\ (p,q))] + [1/(\alpha m + n)][\Delta \text{AC}]$, for a developing country.

The change in aggregate cost = $\Delta \text{AC} = -\left[\sum_{i \in M} C_i(x_{i,c}(m,n)) + \sum_{j \in N} C_j(x_{j,c}(m,n))\right] + \left[\sum_{i \in P} C_i(x_{i,c}(p,q)) + \sum_{j \in Q} C_j(x_{j,c}(p,q)) + \sum_{i \in M \backslash P} C_i(x_{i,n}(p,q)) + \sum_{j \in N \backslash Q} C_j(x_{j,n}(p,q))\right]$.

In addition, because all the monetary transfers must add up to zero, we have $\sum_{i \in P} \tau_{i,c} + \sum_{j \in Q} \tau_{j,c} + \sum_{i \in M \backslash P} \tau_{i,n} + \sum_{i \in N \backslash Q} \tau_{j,n} = 0$. From Sect. 3, we know that the only non-trivial coalition formed is $\{(p,q)\} = \{(3,0)\}$. It follows that the monetary transfers can be

[10] See Finus and Rundshagen (2003b) or Carraro et al. (2006) for details.

simplified as $\sum_{i \in P} \tau_{i,c} + \sum_{i \in M \setminus P} \tau_{i,n} + \sum_{i \in N} \tau_{j,n} = 3\tau_{i,c} + (m-3)\tau_{i,n} + n\tau_{j,n} = 0$. Therefore, the following proposition holds:

Proposition 4.1 *The developed countries inside the small coalition are the net contributors of the monetary transfers, while the developing countries (outside the small coalition) are the net receivers. The developed countries outside the small coalition are also the net contributors if α is slightly larger than* 1.

Therefore, the developed countries that originally form the small coalition have to offer side payments to the developing countries outside the small coalition to induce them to form the grand coalition. When α is slightly larger than 1, the developed countries outside the small stable coalition also have to be the net contributors, although they pay less than the three countries inside the small stable coalition. The focus of our article is on the cases when the other developed countries are also the net contributors.

In addition, recall that for the developed countries, $U_{i,n}$ (3,0) > $U_{i,c}$ (3,0), which means that in the first stage the developed countries outside the small coalition are better off than the developed countries inside the small coalition. We also know that $\tau_{i,c} < \tau_{i,n}$ from Lemma 4.1 (in Appendix). It follows that $Y_{i,c} = W_i + \tau_{i,c} < W_i + \tau_{i,n} = Y_{i,n}$. This means that, with the monetary transfers in the second stage, the developed countries outside the small coalition are even better off than the developed countries originally inside the small coalition. Therefore, there is a *later-mover* advantage for the developed countries both before and after countries engage in monetary transfers to form the grand coalition.[11]

4.2 The subgame perfectness of the expansion path

In this section, we construct a strategy profile or combination of all the countries in both the first stage and the second stage to constitute a subgame perfect Nash equilibrium (SPNE). Because there are many strategies of many countries, there can be a large number of subgames. (Some of them are discussed in Sect. 4.3.) For reasons of simplicity, we focus only on the scenarios that we have discussed in Sects. 3 and 4.

More specifically, because the game is an ECFG with monetary transfers, the countries that intend to form a non-trivial coalition must agree on which countries should be in the coalition. At the same time, if these countries engage in monetary transfers, there are net contributors and net receivers. All the net contributors and net receivers must also agree on the amounts of transfers. As described in Sects. 3 and 4, at the beginning of the first stage, all countries form singletons. Then, three developed countries move first to form a small stable coalition without monetary transfers, which requires the agreements of these three developed countries. In the second stage, all countries engage in monetary transfers to form the grand coalition, which requires a higher degree of consensus among all the countries.

Moreover, in the first stage, for these three developed countries that intend to form a small stable coalition, their strategies are to form this small stable coalition {(3,0)}, with one another without monetary transfers. All the other countries remain "inactive" in the first stage as singletons or, more precisely, as free riders. The strategies of all the other developed and developing countries therefore are to form singletons in the first stage. Given that the small coalition is both internally and externally stable, no country has the incentive to deviate from its chosen strategy given the strategy chosen by all the other

[11] Because of the symmetry assumption for the same type of countries, there is no good reason to justify why some developed countries form the small coalition first if they know they can be better off by letting other developed countries do so. However, this can be justified if all countries are totally heterogeneous.

countries. Hence, the strategy combination of all the countries constitutes a Nash equilibrium in the first stage.

In the second stage, countries engage in monetary transfers to form the grand coalition with monetary transfers, because it is a Pareto improvement for all the countries. However, such an expansion in the second stage from the small stable coalition to the grand coalition with monetary transfers depends on the satisfaction of all the following four conditions: First, the small stable coalition {(3,0)} is already formed in the first stage. Second, all countries must agree to participate in forming the grand coalition. Third, monetary transfers are available as defined in Definition 4.1. Fourth, all countries, including all the net contributors and net receivers, must agree on the amounts of monetary transfers. Otherwise, there is no grand coalition or monetary transfers, and the three developed countries continue to form the same small stable coalition {(3,0)}. Because of this, in the second stage, no country has the incentive to deviate from its strategy, given the strategy chosen by all the other players. Therefore, the strategy combination of all the countries in the second stage also constitutes a Nash equilibrium.

From the backward induction, in order for all countries to form the grand coalition with monetary transfers in the second stage, the three developed countries must form the small coalition $\{(p,q)\} = \{(3,0)\}$ without monetary transfers in the first stage, while all the other developed and developing countries form singletons. Therefore, the expansion path from {(3,0)} to $\{(m,n)\}$ is the only outcome of backward induction. But the complete description of the subgame perfect Nash equilibrium is more complicated, because the strategy combination of all the countries must constitute a Nash equilibrium in every subgame. The details are presented in Appendix.

4.3 Other possible subgames

There are many other possible subgames or paths that will not be realized. For reasons of simplicity, we would discuss two of them and compare them with the subgame that we discussed in Sect. 4.2. More specifically, Path I is for countries to form the grand coalition $\{(m,n)\}$ directly from {(0,0)} without monetary transfers. Path II is for countries to form the grand coalition $\{(m,n)\}$ directly from {(0,0)}, but with monetary transfers. And Path III is from {(0,0)} to {(3,0)} without monetary transfers and then from {(3,0)} to $\{(m,n)\}$ with monetary transfers, which is subgame discussed in Sect. 4.2.

Also, assume that only the final payoffs or utilities matter to the countries and that it is possible for countries to reach the consensus to form the grand coalition at the beginning of the first stage when all countries form singletons, as discussed by Chander and Tulkens (1995). If all the countries can reach such an agreement to form the grand coalition, then the game is over, although it may take a long time for all countries to reach the unanimous agreement. Hence, an important question is whether the developed and developing countries should engage in monetary transfers to form the grand coalition after the small coalition is formed (Path III), or should they do so at the beginning of the first stage before the small coalition is formed (Path II). To answer this question, let $Y_{i,n}\ [(m,n)|(0,0)\rightarrow(m,n)]$ and $Y_{i,c}\ [(m,n)|(3,0)\rightarrow(m,n)]$ represent the payoffs of a developed country, when all countries engage in monetary transfers before and after the formation of the small stable coalition, respectively. Similarly, let $Y_{j,n}\ [(m,n)|(0,0)\rightarrow(m,n)]$ and $Y_{j,n}\ [(m,n)|(3,0)\rightarrow(m,n)]$ represent the payoffs of a developing country, when all countries engage in monetary transfers before and after the formation of the small stable coalition, respectively. Then, consistent with Proposition 4.1, the following proposition holds:

Proposition 4.2 *It is beneficial for the developed countries to engage in utility transfers before the small coalition is formed, whereas it is beneficial for the developing countries to engage in utility transfers after the developed countries form the small coalition.*[12]

Note that to form the grand coalition requires the agreements and the participations of all the countries. Although it is beneficial for the developed countries to form the grand coalition before the small coalition is formed (Path II), the developing countries will not participate in the treaty because they can benefit more when all countries engage in monetary transfers after the small coalition is formed (Path III). At the same time, because of the late-mover advantage, other developed countries may not want to participate in forming the grand coalition either. Therefore, the grand coalition cannot be formed, and Path II cannot be realized. Therefore, only Path III will be realized—the three developed countries should move first to form the small stable coalition in the first stage. Then, all the countries engage in utility transfers to form the grand coalition in the second stage.

Among the three paths, Path I is the first best outcome for the developed countries because they do not have to pay the developing countries, but it is the third best outcome for the developing countries. At the same time, because countries have the incentive to be free riders, the grand coalition may not be stable even if it can be formed. This may explain why in the early stages of the Kyoto Protocol without technology or monetary transfers, countries had difficulties in forming the grand coalition on Path I. For Path II, it is the second best outcome for all the developed and developing countries. For Path III, it would be the third best outcome for the three developed countries in the small coalition, but the first best outcome for the developing countries because the developing countries receive the highest side payments from the developed countries.

Nevertheless, consensus building takes time, and it may take a long time for all the countries to reach an agreement on Path I or Path II even if it is possible. Given the uncertainties of future damages in the case of global warming, the short-term benefit may be important as well. Therefore, the best policy for the three developed countries may still be Path III—to form a small coalition of reducing emissions among themselves without monetary transfers, and without the participation of all the developed countries, such as the USA. The developing countries will start reducing their emissions later on in the future, when the monetary and technology transfers become available.

5 Extensions

From the preceding propositions, results, and discussions, there are at least three important scenarios that are worth exploring for future extensions of this article. They are alternative punishments for deviating (developed) countries, repeated applications of monetary transfers to form the grand coalition, and the formation of multiple small coalitions under different rules of the game.

5.1 Alternative punishment for deviating actions

When countries form the grand coalition based on the consensus of all countries, a deviating country will trigger the breakup of the grand coalition back to the small stable

[12] Chander and Tulkens (1995) have a similar perspective in their proof on p. 288.

coalition. If the possible deviating country or countries are the developing countries, the problem is easier to resolve. Given that a deviating developing country loses the side payment, the developed countries can raise the opportunity cost of deviation by increasing the payment (i.e., the carrot) and/or imposing some penalty (i.e., the stick) to decrease the free-riding incentive of the developing countries or even to prevent developing countries from deviating. However, if a developed country, such as the USA, deviates or withdraws from a treaty, it is more difficult to use the "carrot-and-stick" method. This is because developed countries are the net contributors of the monetary transfers, particularly the three developed countries that form the small coalition in the first stage. Hence, deviation from the grand coalition means that they do not have to pay the developing countries to join the IEA to form the grand coalition. In addition, while developed countries, such as the USA, can impose a trade sanction on deviating developing countries, which country can impose such a penalty on the USA? This is one important question that is worth further research.

5.2 Repeated applications of monetary transfers via a sharing rule

Another way to expand the small coalition to the grand coalition is the repeated applications of the monetary transfers via a sharing rule, one country or some countries at a time, as proposed by Weikard (2005). Recall that to engage in utility transfers via the sharing rule is Pareto improving such that no countries are worse off and at least one country is strictly better off than the status quo. Therefore, any coalition smaller than the grand coalition is not externally stable, as long as there are efficiency gains for countries to engage in utility transfers to form a larger coalition. At the same time, because a deviating country will lose its side payment, no country should have an incentive to do so, and thus the coalition formed via the sharing rule is internally stable. Therefore, the repeated applications of the sharing rule can expand a small stable coalition to a larger coalition that is internally stable until the grand coalition is formed. The follow-up question is what the optimal path is from the small stable coalition to the grand coalition. Such an issue can be explored in future research.

5.3 Other rules of endogenous coalition formation

There are other rules of games in endogenous coalition formation. In some of the games, there can be more than one non-singleton coalition, with or without utility transfers. From a modeling perspective, it may be more efficient for countries to form several small stable coalitions in the first stage because of a higher level of public good provided, although they have to pay the outsiders more side payments to form the grand coalition in the second stage. In other words, instead of forming one non-trivial coalition, should countries form several small coalitions in the first stage? Such a coalition formation may be important if the short run benefit matters more than the long run benefit of forming the grand coalition. This is another important direction worth exploring.

6 Conclusion

This research builds a two-stage ECFG in which countries have the option to engage in utility transfers to form a coalition in both stages. In the first stage of an IEA formation process, because most countries have incentives to be free riders, there are only a few participating countries. The developed countries have to decide whether or not to include a developing country in the non-trivial coalition. Our results show that engaging in monetary transfers will accomplish little because only one developing country will participate in forming a non-trivial coalition with one developed country. Instead, it is more efficient for all the countries if three developed countries form a stable non-trivial coalition with zero or no utility transfers. It is also easier for participating countries to reach an agreement, because it requires the consensus of only three developed countries.

In the second stage, countries expand the small stable coalition to the grand coalition by engaging in monetary transfers. The three developed countries originally in the small stable coalition have to pay the largest amounts of monetary transfers to induce the developing countries to join the grand coalition. The policy implication is that developing countries should not participate in the treaty in the first stage of an IEA formation process. Instead, they should join the grand coalition as late as possible or at least wait until monetary transfers become available. The other developed countries that have chosen to form singletons in the first stage probably have to pay developing countries to participate in forming the grand coalition in the second stage. But these developed countries pay less than those that have formed the small stable coalition in the first stage. This may explain why developed countries, such as the USA, also have incentives to be free riders in the early stage of an IEA formation process. We have defined the strategy profile or combination of all the countries that constitutes the subgame perfect Nash equilibrium for the two-stage ECFG of an IEA formation process. We have also shown that the expansion path from {(0,0)} to {(3,0)} without monetary transfers and then from {(3,0)} to {(*m*,*n*)} with monetary transfers is the only outcome of backward induction.

We have shown that while the developed countries prefer to form the grand coalition directly from the situations when every country forms a singleton, the developing countries prefer that three developed countries form a stable non-trivial coalition first. It is the first best outcome for the developing countries, although it is a third best outcome for the developed countries originally in the small stable coalition. However, it is still better than the situation when three developed countries form the stable non-trivial coalition, or when every country continues to form singletons.

Finally, it is possible, but probably unlikely, that a global consensus can be achieved at the beginning of an IEA formation process, in particular if there are high levels of uncertainties involved. Therefore, the analysis provided in this article can explain to some extent the historical development of the Montreal Protocol, as well as the question of why countries had difficulties in forming the grand coalition in the Kyoto Protocol in the early stages of an IEA formation process.

Appendix

Lemma 2.1 *$\tau_i = 0$ and $\tau_j = 0$ occur when all the participating countries are symmetric.*

Proof When only p developed countries employ the sharing rule, $\tau_i = -[W_i - V_i] + [\alpha/(\alpha p)][\sum_{i\in P}(W_i - V_i)] = 0$. When only q developing countries employ the sharing rule, $\tau_j = -[W_j - V_j] + (1/q)[\sum_{j\in Q}(W_i - V_i)] = 0$. □

Lemma 3.1 *$\tau_i < 0$ and $\tau_j > 0$.*

Proof

$$\tau_i = -[W_i(p,q) - V_i] + [\alpha/(\alpha p + q)]\left[W - \left(\sum_{i\in P} V_i + \sum_{j\in Q} V_j\right)\right]$$
$$= [W_i(p,q) - V_i][(-q)/(\alpha p + q)] + [(\alpha q)/(\alpha p + q)][W_j(p,q) - V_j]$$
$$= [q/(\alpha p + q)]\{-[W_i(p,q) - V_i] + (\alpha)[W_j(p,q) - V_j]\}$$
$$= [q/(\alpha p + q)](1/2)(1-\alpha)[(\alpha p + q)^2 + \alpha] < 0.$$

$$\tau_j = -[W_j(p,q) - V_j] + [1/(\alpha p + q)]\left[W - \left(\sum_{i\in P} V_i + \sum_{j\in Q} V_j\right)\right]$$
$$= [(-p)/(\alpha p + q)]\{-[W_i(p,q) - V_i] + (\alpha)[W_j(p,q) - V_j]\} > 0.$$ □

Proof of Proposition 3.1 This follows Lemma 3.1. □

Lemma 3.2

$$Y_i(p,q) = (\alpha)[(\alpha)(m-p) + (n-q) + (1/2)(\alpha p + q)(p+q) + (1/2)(1-\alpha)(q)/(\alpha p + q)].$$

$$Y_j(p,q) = (\alpha)(m-p) + (n-q) + (1/2)(\alpha p + q)(p+q) + (1/2)(\alpha p)(\alpha - 1)/(\alpha p + q).$$

$$\frac{\partial Y_i(p,q)}{\partial p} > 0, \frac{\partial Y_i(p,q)}{\partial q} > 0, \frac{\partial Y_j(p,q)}{\partial p} > 0, \frac{\partial Y_j(p,q)}{\partial q} > 0.$$

Proof

$$U_{i,c}(p,q) = \alpha[\alpha(m-p) + (n-q) + (\alpha p + q)(p+q)] - (1/2)(\alpha p + q)^2, \quad \text{if } p \ge 1.$$
$$U_{i,n}(p,q) = \alpha[\alpha(m-p) + (n-q) + (\alpha p + q)(p+q)] - (1/2)(\alpha^2).$$
$$U_{j,c}(p,q) = [\alpha(m-p) + (n-q) + (\alpha p + q)(p+q)] - (1/2)(\alpha p + q)^2, \quad \text{if } q \ge 1.$$
$$U_{j,n}(p,q) = [\alpha(m-p) + (n-q) + (\alpha p + q)(p+q)] - (1/2).$$

$$Y_i(p,q) = W_i - [W_i - V_i] + [\alpha/(\alpha p + q)]\left[W - \left(\sum_{i\in P} V_i + \sum_{j\in Q} V_j\right)\right]$$
$$= U_{i,n}(0,0) + [\alpha/(\alpha p + q)]\left[\sum_{i\in P}(U_{i,c}(p,q) - U_{i,n}(0,0)) + \sum_{j\in Q}(U_{j,c}(p,q) - U_{j,n}(0,0))\right]$$
$$= (\alpha)[(\alpha)(m-p) + (n-q) + (1/2)(\alpha p + q)(p+q) + (1/2)(1-\alpha)(q)/(\alpha p + q)].$$

$$Y_j(p,q) = W_j - [W_j - V_j] + [1/(\alpha p + q)]\left[W - \left(\sum_{i\in P} V_i + \sum_{j\in Q} V_j\right)\right]$$
$$= U_{j,n}(0,0) + [1/(\alpha p + q)]\left[\sum_{i\in P}(U_{i,c}(p,q) - U_{i,n}(0,0)) + \sum_{j\in Q}(U_{j,c}(p,q) - U_{j,n}(0,0))\right]$$
$$= (\alpha)(m-p) + (n-q) + (1/2)(\alpha p + q)(p+q) + (1/2)(\alpha p)(\alpha - 1)/(\alpha p + q).$$

$$\frac{\partial Y_i(p,q)}{\partial p} = (\alpha/2)[(\alpha-1)(2p) + (\alpha+1)(q) + (\alpha)(\alpha-1)(q)/(\alpha p+q)^2] > 0.$$

$$\frac{\partial Y_i(p,q)}{\partial q} = (\alpha/2)[(p+q-2) + (\alpha p+q) - (\alpha^2 p+q)/(\alpha p+q)^2] > 0.$$

$$\frac{\partial Y_j(p,q)}{\partial p} = (1/2)[(2\alpha)(p-1) + (\alpha+1)(q) + (\alpha)(\alpha-1)(q)/(\alpha p+q)^2] > 0.$$

$$\frac{\partial Y_j(p,q)}{\partial q} = (1/2)[(p+q-2) + (\alpha p+q) + (\alpha p)(1-\alpha)/(\alpha p+q)^2] > 0.$$ □

Lemma 3.3 *$F(p,q) - \alpha G(p,q) \geq 0$, where $F(p,q) \equiv Y_i(p,q) - U_{i,n}(p-1,q)$ and $G(p,q) \equiv Y_j(p,q) - U_{j,n}(p,q-1)$.*

Proof

$F(p,q) = (\alpha)[(-1/2)(\alpha p + q)(p + q) + (2\alpha)(p - 1) + (\alpha + 1)(q) + (\alpha^2 p + q)/(2\alpha p + 2q)]$. $G(p,q) = (-1/2)(\alpha p + q)(p + q) + 2(q - 1) + (\alpha + 1)(p) + (\alpha^2 p + q)/(2\alpha p + 2q)$. Hence, $F(p,q) - \alpha G(p,q) = (\alpha)(\alpha - 1)(p + q - 2) \geq 0$, since $\alpha > 1$ and $p + q \geq 2$. □

Proof of Proposition 3.2 The satisfaction of the internal stability conditions means that $F(p,q) \geq 0$ and $G(p,q) \geq 0$. From Lemma 3.3, $G(p,q) \geq 0$ implies that $F(p,q) \geq 0$. Hence, if a developing country does not have an incentive to deviate, neither does a developed country. □

Lemma 3.4 $G(1,1) > 0, G(1,2) < 0, G(2,1) < 0, \frac{\partial G(p,q)}{\partial q} < 0, \frac{\partial G(p,q)}{\partial p} < 0, F(1,1) > 0, F(1,2) > 0, F(2,1) > 0, F(3,1) < 0, \frac{\partial F(p,q)}{\partial q} < 0$ *if* $p \geq 2$; $\frac{\partial F(p,q)}{\partial p} < 0$ *if* $p \geq 3$.

Proof

$$G(1,1) = (\alpha^2+1)/(2\alpha+2) > 0.$$
$$G(1,2) = (1-\alpha)/(\alpha+2) < 0.$$
$$G(2,1) = (1-\alpha)(1+\alpha)/(2\alpha+1) < 0.$$

$$\frac{\partial G(p,q)}{\partial q} = (-1/2)[(p+q-2) + (\alpha p+q-2) + (\alpha-1)(\alpha p)/(\alpha p+q)^2] < 0.$$

$$\frac{\partial G(p,q)}{\partial p} = (-1/2)[(\alpha+1)(p+q-2) + (\alpha-1)(p) + (1-\alpha)(\alpha p)/(\alpha p+q)^2] < 0.$$

From Lemma 3.3, $F(p,q) = \alpha G(p,q) + (\alpha)(\alpha - 1)(p + q - 2)$. Hence, we have $F(1,1) = \alpha G(1,1) > 0$.

$$F(2,1) = (\alpha^2)(\alpha-1)/(2\alpha+1) > 0.$$
$$F(1,2) = (\alpha)(\alpha+1)(\alpha-1)/(\alpha+2) > 0.$$
$$F(3,1) = (-\alpha)(3\alpha^2+8\alpha+2)/(6\alpha+2) < 0.$$

$$\frac{\partial F(p,q)}{\partial q} = (\alpha)[\frac{\partial G(p,q)}{\partial q} + (\alpha - 1)]$$
$$= (-1/2)(\alpha)[2(p+q-2) + (\alpha-1)(p-2) + \alpha p/(\alpha p + q)^2] < 0 \quad \text{if } p \geq 2.$$

$$\frac{\partial F(p,q)}{\partial p} = (\alpha)[\frac{\partial G(p,q)}{\partial p} + (\alpha - 1)$$
$$= (-1/2)(\alpha)[(\alpha+1)(p+q-2) + (\alpha-1)(p-2) - (\alpha p)/(\alpha p + q)^2] < 0 \quad \text{if } p \geq 3.$$

□

Proof of Proposition 3.3 The combination of $G(1,2) < 0$ and $\frac{\partial G(p,q)}{\partial q} < 0$ means that when $p = 1$, and $q \geq 2$, it is beneficial for a developing country to deviate from the non-trivial coalition. Similarly, the combination of $G(2,1) < 0$ and $\frac{\partial G(p,q)}{\partial p} < 0$ means that when $p \geq 2$ and $q = 1$, the developing country should also deviate from the coalition. Given that $G(1,1) > 0$ and $F(1,1) \geq \alpha G(1,1) > 0$, both the internal stability conditions for the developed and developing countries are satisfied when $(p,q) = (1,1)$. Therefore, the non-trivial coalition, $\{(p,q)\} = \{(1,1)\}$, is internally stable for both the developed country and the developing country inside the coalition.

Also, the combination of $F(2,1) > 0$, $F(3,1) < 0$, and $\frac{\partial F(p,q)}{\partial p} < 0$ if $p \geq 3$ means that, when $q = 1$ and $p = 2$, the developed countries do not have an incentive to deviate from the non-trivial coalition. However, when $q = 1$ and $p \geq 3$, even the developed countries have an incentive to deviate from the coalition to become free riders. Therefore, both $\{(p,q)\} = \{(1,1)\}$ and $\{(p,q)\} = \{(2,1)\}$ are internally stable for the developed countries. □

Lemma 3.5 *$E(1,1) < 0, \frac{\partial E(p,q)}{\partial q} < 0, \frac{\partial E(p,q)}{\partial p} < 0, D(1,1) > 0, D(2,1) < 0$, where $D(p,q) \equiv Y_i(p+1,q) - U_{i,n}(p,q)$, and $E(p,q) \equiv Y_j(p,q+1) - U_{j,n}(p,q)$.*

Proof

$$E(1,1) = G(1,2) = (1-\alpha)/(\alpha+2) < 0.$$
$$D(1,1) = F(2,1) = (\alpha^2)(\alpha-1)/(2\alpha+1) > 0.$$
$$D(2,1) = F(3,1) = (-\alpha)(3\alpha^2 + 8\alpha + 2)/(6\alpha+2) < 0.$$
$$G(2,1) = (1-\alpha)(1+\alpha)/(2\alpha+1) < 0.$$

$$\frac{\partial E(p,q)}{\partial q} = (-1/2)[(p+q-1) + (\alpha p + q - 1) + (\alpha-1)(\alpha p)/(\alpha p + q)^2] < 0.$$

$$\frac{\partial E(p,q)}{\partial p} = (-1/2)[(\alpha+1)(p+q-1) + (\alpha-1)(p) + (1-\alpha)(\alpha)(q+1)/$$
$$(\alpha p + q + 1)^2] < 0.$$

□

Proof of Proposition 3.4 The satisfaction of the external stability condition $\{(p,q)\}$ means that $D(p,q) < 0$ and $E(p,q) < 0$. The combination of $E(1,1) < 0$, $\frac{\partial E(p,q)}{\partial q} < 0$, and $\frac{\partial E(p,q)}{\partial p} < 0$ shows that the coalition $\{(p,q)\} = \{(1,1)\}$ is externally stable for the developing countries, as well as for any other coalitions larger than $\{(p,q)\} = \{(1,1)\}$. This means that no developing countries should have the incentive to participate in any coalition larger than or equal to $\{(p,q)\} = \{(1,1)\}$. For the developed countries, the combination of $D(1,1) > 0$

and $D(2,1) < 0$ means that the coalition $\{(p,q)\} = \{(2,1)\}$ is externally stable for the developed countries, while $\{(p,q)\} = \{(1,1)\}$ is not. □

Proof of Proposition 3.5

$$Y_{i,c}(1,1) - U_{i,c}(3,0) = (\alpha)(1-\alpha)/(2\alpha+2) - (3/2)(\alpha^2) < 0.$$
$$Y_{j,c}(1,1) - U_{j,n}(3,0) = (\alpha)(\alpha-1)/(2\alpha+2) - (6\alpha-(1/2)) = (-1)(11\alpha^2+12\alpha-1)/(2\alpha+2) < 0.$$

This means that both the developed country and the developing country forming the non-trivial coalition are better off when $\{(p,q)\} = \{(3,0)\}$ instead of $\{(p,q)\} = \{(1,1)\}$. Also, $Y_{i,n}$ (1,1) − $U_{i,c}$ (3,0) = $(\alpha)(1-\alpha) < 0$. This means that two more developed countries originally outside the coalition $\{(1,1)\}$ also benefit from forming the coalition $\{(3,0)\}$. Therefore, it is more efficient for three developed countries to form the stable coalition $\{(p,q)\} = \{(3,0)\}$ than $\{(p,q)\} = \{(1,1)\}$. □

Lemma 4.1 $\tau_{i,c} < \tau_{i,n} < \tau_{j,n}$.

Proof

$\tau_{i,c} - \tau_{i,n} = U_{i,c} - U_{i,n} = (1/2)[(\alpha)(1+p)+q][(\alpha)(1-p)-q] = (-4)(\alpha^2) < 0$, since $(p,q) = (3,0)$. Therefore, $\tau_{i,c} < \tau_{i,n}$. In addition, from Definition 4.2, we have

$$\begin{aligned}\tau_{i,n} - \tau_{j,n} = {} & [C_i(x_{i,c}(m,n)) - C_i(x_{i,n}(p,q))] - [\alpha/(\alpha m+n)][\Delta\text{AC}] - [C_j(x_{j,c}(m,n)) \\ & - C_j(x_{j,n}(p,q))] + [1/(\alpha m+n)][\Delta\text{AC}], \text{ where } \Delta\text{AC} = \Bigg[\sum_{i\in M} C_i(x_{i,c}(m,n)) \\ & + \sum_{j\in N} C_j(x_{j,c}(m,n))\Bigg] - \Bigg[\sum_{i\in P} C_i(x_{i,c}(p,q)) + \sum_{j\in Q} C_j(x_{j,c}(p,q)) \\ & + \sum_{i\in M\backslash P} C_i(x_{i,n}(p,q)) + \sum_{j\in N\backslash Q} C_j(x_{j,n}(p,q))\Bigg] \\ = {} & [C_i(x_{i,c}(m,n)) - C_i(x_{i,n}(p,q))] - [C_j(x_{j,c}(m,n)) - C_j(x_{j,n}(p,q)] \\ & + [(1-\alpha)/(\alpha m+n)][\Delta\text{AC}] \\ = {} & (1/2)(1-\alpha)(1+\alpha) + [(1-\alpha)/(\alpha m+n)](1/2)[(m)(\alpha m+n)^2 \\ & + (n)(\alpha m+n)^2 - (p)(\alpha p+q)^2 - (m-p)(\alpha^2) - (1/2)(n)] < 0, \text{ since } 1<\alpha.\end{aligned}$$

Therefore, we have $\tau_{i,n} < \tau_{j,n}$. □

Proof of Proposition 4.1 From Lemma 4.1, we have $\tau_{i,c} < \tau_{i,n} < \tau_{j,n}$. Since $\sum_{i\in P}\tau_{i,c} + \sum_{i\in M\backslash P}\tau_{i,n} + \sum_{j\in N}\tau_{j,n} = p\tau_{i,c} + (m-p)\tau_{i,n} + n\tau_{j,n} = 0$ and $n > m > p = 3$, it follows that $\tau_{i,c} < 0$ and $\tau_{j,n} > 0$. Also, because $\tau_{i,n} = (1/2)[1/(\alpha m+n)][(n)(1-\alpha)(\alpha m+n)^2 + (\alpha n)(1-\alpha) + 24\alpha^3]$, the sign of $\tau_{i,n}$ depends on the sign of $[(n)(1-\alpha)(\alpha m+n)^2 + (\alpha n)(1-\alpha) + 24\alpha^3]$. Let $f(m,n,\alpha) \equiv (n)(1-\alpha)(\alpha m+n)^2 + (\alpha n)(1-\alpha) + 24\alpha^3$. Recall that $n > m > 3$, it follows that $(m,n) \geq (4,5)$ and that $f(4,5,1.1) = -12.79 < 0$. In addition,

$$\frac{\partial f(m,n,\alpha)}{\partial m} = (2)(1-\alpha)(\alpha n)(\alpha m+n) < 0.$$

$$\frac{\partial f(m,n,\alpha)}{\partial n} = (1-\alpha)[(\alpha m+n)^2 + (2n)(\alpha m+n) + \alpha] < 0.$$

$\frac{\partial f(m,n,\alpha)}{\partial \alpha} = (-n)(\alpha m+n)^2 + (1-\alpha)(2mn)(\alpha m+n) + (n)(1-2\alpha) + 72\alpha^2 < 0$ when $(m,n) = (4,5)$. With $\frac{\partial[\frac{\partial f(m,n,\alpha)}{\partial \alpha}]}{\partial m} < 0$, and $\frac{\partial[\frac{\partial f(m,n,\alpha)}{\partial \alpha}]}{\partial n} < 0$, it follows that $\frac{\partial f(m,n,\alpha)}{\partial \alpha} < 0$ for all $(m,n) \geq (4,5)$. Hence, $f(m,n,\alpha) < 0$ for all $(m,n) \geq (4,5)$ when α is slightly larger than 1. □

Lemma 4.2 $Y_{i,n}$ $[(m,n)|(0,0) \rightarrow (m,n)] > Y_{i,c}$ $[(m,n)|(3,0) \rightarrow (m,n)]$ and $Y_{j,n}$ $[(m,n)|(0,0) \rightarrow (m,n)] < Y_{j,n}$ $[(m,n)|(3,0) \rightarrow (m,n)]$.

Proof

(1)

$$\begin{aligned}
&Y_{i,n}[(m,n)|(0,0) \rightarrow (m,n)] - Y_{i,c}[(m,n)|(3,0) \rightarrow (m,n)] \\
&= U_{i,n}(0,0) + [\alpha/(\alpha m+n)]\{W - [mU_{i,n}(0,0) + nU_{j,n}(0,0)]\} - U_{i,c}(3,0) - \\
&\quad [\alpha/(\alpha m+n)]\{W - [3U_{i,c}(3,0) + (m-3)U_{i,n}(3,0) + nU_{j,n}(3,0)]\} \\
&= [U_{i,n}(0,0) - U_{i,c}(3,0)][1 - \alpha/(\alpha m+n)] + [\alpha/(\alpha m+n)][(m-3)(U_{i,n}(3,0) - U_{i.c}(3,0)) + \\
&\quad n(U_{j,n}(3,0) - U_{j,n}(0,0))] \\
&= (\alpha^2)[4n + (4\alpha)(m-3)]/(\alpha m+n) > 0, \text{ since } m > 3.
\end{aligned}$$

(2)

$$\begin{aligned}
&Y_{j,n}\ [(m,n)|(0,0) \rightarrow (m,n)] - Y_{j,n}[(m,n)|(3,0) \rightarrow (m,n)] \\
&= U_{j,n}(0,0) + [1/(\alpha m+n)]\{W - [mU_{i,n}(0,0) + nU_{j,n}(0,0)]\} - U_{j,n}(3,0) - \\
&\quad [1/(\alpha m+n)]\{W - [3U_{i,c}(3,0) + (m-3)U_{i,n}(3,0) + nU_{j,n}(3,0)]\} \\
&= [U_{j,n}(0,0) - U_{j,n}(3,0)] + [1/(\alpha m+n)][(3)(U_{i,c}(3,0) - U_{i.n}(0,0)) + \\
&\quad (m-3)(U_{i,n}(3,0) - U_{i.n}(0,0)) + (n)(U_{j,n}(3,0) - U_{j,n}(0,0))] = (-12)(\alpha^2)/(\alpha m+n) < 0.
\end{aligned}$$ □

Proof of Proposition 4.2 This directly follows Lemma 4.2. □

Subgame perfect Nash equilibrium (SPNE)

When countries engage in monetary transfers, there are other possibilities than those specified in Definition 4.1, including no monetary transfers. Hence, forming the same coalition with different monetary transfers should constitute a different subgame. Because of this, each country actually has a very large or infinite number of strategies, and there are also many subgames for the whole game. Therefore, as a SPNE is a strategy combination that constitutes a Nash equilibrium in every subgame, the strategy combination must also take all the possible monetary transfers into consideration, at least in the second stage.

In the first stage, all the three developed countries that intend to form the small stable coalition have the following strategy:

$S_{1,i}$ = Form the small stable coalition {(3,0)} with two other developed countries with no monetary transfers.

All the other $(m-3)$ developed countries have the following strategy:

$S_{1,i}$ = Form a singleton.

All n developing countries have the following strategy:

$S_{1,j}$ = Form a singleton.

In the second stage, for the three developed countries that have already formed the small stable coalition $\{(3,0)\}$ in the first stage, their strategies are as follows:

$S_{2,i}$ = If the following three conditions are satisfied, then offer side payments to the developing countries to form the grand coalition:

(A) All countries must agree to participate in forming the grand coalition.
(B) The side payments paid by each of these three developed countries cannot exceed the amounts specified in Definition 4.1.
(C) All countries must agree on the amounts of monetary transfers.
Otherwise, $S_{2,i}$ = Continue to form $\{(p,q)\} = \{(3,0)\}$ without monetary transfers.

For all the other $(m - 3)$ developed countries that form singletons in the first stage, they have the following strategies in the second stage:

$S_{2,i}$ = If all the following four conditions are satisfied, then offer side payments to the developing countries to form the grand coalition:

(A) All countries must agree to participate in forming the grand coalition.
(B) The small stable coalition $\{(3,0)\}$ is already formed in the first stage by the other three developed countries.
(C) The side payments paid by each of these $(m - 3)$ developed countries cannot exceed the amounts specified in Definition 4.1.
(D) All countries must agree on the amounts of monetary transfers.
Otherwise, $S_{2,i}$ = Continue to form singletons.

All the n developing countries have the following strategies in the second stage:

$S_{2,j}$ = If all the following four conditions are satisfied, form the grand coalition:

(A) All countries must participate in forming the grand coalition.
(B) The three developed countries have formed a small stable coalition $\{(3,0)\}$ in the first stage.
(C) The side payments received by each developing country cannot be less than the amounts specified in Definition 4.1.
(D) All countries must agree on the amounts of monetary transfers.
Otherwise, $S_{2,j}$ = Continue to form only singletons.

Therefore, no country can be better off by unilaterally deviating from its strategy, given the strategy chosen by all the other countries. Let S_1 be the strategy combination of all the countries in stage 1, and S_2 be the strategy combination of all the countries in stage 2. Then S_1 is a Nash equilibrium in the first stage because no country in the first stage can be strictly better off by changing its strategy given the strategy chosen by all the other countries. For the same reasoning, in the second stage, S_2 is a Nash equilibrium in stage 2. If any country deviates from its strategy, then the grand coalition cannot be formed, and the deviating country in stage 2, along with all the countries, is strictly worse off. Hence, countries will form the grand coalition in the second stage with monetary transfers. Because of this, from the backward induction, the developed countries in the first stage must move fist to form the small stable coalition $\{(3,0)\}$. If any other subgame is realized in the first stage, the grand coalition cannot be formed in the second stage, and the three developed countries will continue to form the small stable coalition {(3,0)}. Therefore, the expansion path from {(3,0)} to {(*m*,*n*)} with monetary transfers defined in Definition 4.1 is an outcome of backward induction. And the strategy combination for the whole game, $S = (S_1, S_2)$ is also a Nash equilibrium for the whole game, although S_2 given any (other) S_1 is also a Nash equilibrium. Therefore, the strategy combination of all the countries constitutes a Nash equilibrium in every subgame and thus is a subgame perfect Nash equilibrium.

References

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers, 46*, 878–894.

Barrett, S. (1997). Towards a theory of international environmental cooperation. In C. Carraro & D. Siniscalco (Eds.), *New directions in the economic theory of the environment* (pp. 239–280). Cambridge: Cambridge University Press.

Bloch, F. (1995). Endogenous structures of association in oligopolies. *RAND Journal of Economics, 26*(3), 537–556.

Bloch, F., (1996). Sequential formation of coalitions in games with externalities and fixed payoff division. *Games and Economic Behavior, 14*, 90–123.

Bloch, F. (1997). Non-cooperative models of coalition formation in games with spillovers. In C. Carraro & D. Siniscalco (Eds.), *New directions in the economic theory of the environment* (pp. 311–352). Cambridge: Cambridge University Press.

Botteon, M., & Carraro, C. (1997). Environmental coalitions with heterogeneous countries: Burden-sharing and carbon leakage. In A. Ulph (Ed.), *Environmental policy, international agreements, and international trade* (Chap. 3, pp. 38–65). Oxford University Press.

Brèchet, T., Gerard, F., & Tulkens, H. (2007). Climate coalitions: A theoretical and computational appraisal. CORE Discussion Paper, 2007/3.

Caparrós, A., Hammoudi, A., & Tazdaït, T. (2004). On coalition formation with heterogeneous agents. Nota Di Lavoro, 70.2004. *Economics Energy Environment*. Fondazione Eni Enrico Mattei.

Carraro, C., Eyckmans, J., & Finus, M. (2006). Optimal transfers and participation decisions in international environmental agreements. *Review of International Organizations, 1*(4), 379–396.

Carraro, C., & Siniscalco, D. (1991). The International protection of the environment—Voluntary agreements among sovereign countries. Nota Di Lavoro, 1.91. *Economics Energy Environment*. Fondazione Eni Enrico Mattei.

Carraro, C., & Siniscalco, D. (1993). Strategies for the international protection of the environment. *Journal of Pubic Economics, 52*(3), 309–328.

Chander, P. (2007). The γ-core and coalition formation. *International Journal of Game Theory, 35*(4), 539–556.

Chander, P., & Tulkens, H. (1992). Theoretical foundations of negotiations and cost sharing in transfrontier pollution problems. *European Economic Review, 36*(2–3), 388–399.

Chander, P., & Tulkens, H. (1995). A core-theoretical solution for the design of cooperative agreements on transfrontier pollution. *International Tax and Public Finance, 2*(2), 279–293.

Chander, P., & Tulkens, H. (1997). The core of an economy with multilateral environmental externalities. *International Journal of Game Theory, 26*(3), 379–401.

Chander, P., & Tulkens, H. (2001). Limits to climate change. Nota Di Lavoro, 42.2001. *Economics Energy Environment*. Fondazione Eni Enrico Mattei.

Chander, P., & Tulkens, H. (2006). Cooperation, stability, and self-enforcement in international environmental agreements: A conceptual discussion. Nota Di Lavoro, 34.2006. *Economics Energy Environment*. Fondazione Eni Enrico Mattei.

Chander, P., Tulkens, H., Van Ypersele, J.-P., & Willems, S. (1999). The Kyoto protocol: An economic and game theoretic interpretation. Nota Di Lavoro, 72.1999. *Economics Energy Environment*, Fondazione Eni Enrico Mattei.

Chou, P. B., & Shogren, J. F. (2008). *A non-cooperative generalized sharing rule for endogenous coalition formation of heterogeneous countries, as applied to international environmental agreements*, Working Paper.

Chou, P. B., & Sylla, C. (2006). *Endogenous coalition formation with non-transferable utility in international collaboration for global environmental changes*, Working Paper.

d'Aspremont, C. A., Jacqemin, A., Gabszewicz, J. J., & Weymark, J. (1983). On the stability of collusive price leadership. *Canadian Journal of Economics, 16*, 17–25.

Eyckmans, J., & Finus, M. (2004). An almost ideal sharing scheme for coalition games with externalities. Nota Di Lavoro, 155.2004. *Economics Energy Environment*. Fondazione Eni Enrico Mattei.

Eyckmans, J., & Finus, M. (2006). New roads to international environmental agreements: The case of global warming. *Environmental Economics and Policy Studies, 7*(4), 391–414.

Eyckmans, J., & Tulkens, H. (2003). Simulating coalitionally stable burden sharing agreements for the climate change problem. *Resource and Energy Economics, 25*, 299–327.

Figuières, C., & Magali, V. (2003). On the core of an economy with multilateral and multidimensional environmental externalities. *Economics Bulletin, 3*(3), 1–10.

Finus (2003a). Stability and design of international environmental agreements: The case of transboundary pollution. In H. Folmer & T. Tietenberg (Eds.), The international yearbook of environmental and resource economics 2003/2004—A survey of current issues. Cheltenham, UK: Edward Elgar, pp. 82–158.

Finus (2003b) New development in coalition theory. In L. Marsiliani, M. Rauscher, & C. Withagen (Eds.), Environmental policy in an international perspective (pp. 19–49). The Netherlands: Kluwer Academic Publishers.

Finus, M., & Rundshagen, B. (2003a). A non-cooperative foundation of core stability in positive externality NTU-coalitional games. Nota Di Lavoro 31.2003. *Economics Energy Environment*. Fondazione Eni Enrico Mattei.

Finus, M., & Rundshagen, B. (2003b). How the rules of coalition formation affect stability of international environmental agreements. Nota Di Lavoro 62.2003. *Economics Energy Environment*. Fondazione Eni Enrico Mattei.

Finus, M., & Rundshagen, B. (2003c). Endogenous coalition formation in global pollution control: A partition function approach. In C. Carraro (Ed.), *Endogenous formation of economic coalitions* (Chap. 6, pp. 199–243). Cheltenham, UK: Edward Elgar.

Germain, M., Toint, P. L., & Tulkens, H. (1997). Financial transfer to ensure cooperative international optimality in stock pollutant abatement. CORE Discussion Paper, 9701.

Germain, M., Toint, P., Tulkens, H., & De Zeeuw, A. J. (2003). Transfers to sustain dynamic core-theoretical cooperation in international stock pollutant control. *Journal of Economic Dynamics and Control, 28*(1), 79–99.

Germain, M., & van Ypersele, J. -P. (1999). Financial transfers to sustain international cooperation in the climate change framework. CORE Discussion Paper, 9936.

Hart, S., & Kurz, M. (1983). Endogenous formation of coalitions. *Econometrica, 51*, 1047–1064.

Na, S.-L., & Shin, H. S. (1998). International environmental agreements under uncertainty. *Oxford Economic Papers, 50*(2), 173–185.

Shogren, J. F. (1999). *The benefits and costs of the Kyoto protocol*. Washington DC: AEI Press.

Tulkens, H. (1998). Cooperation vs. free riding in international environmental affairs: Two approaches. In N. Hanley & H. Folmer (Eds.), *Game theory and the environment* (Chap. 2, pp. 30–44). Cheltenham: Edward Elgar, et al.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*, Princeton: Princeton University Press.

Weikard, H. -P. (2005). Cartel stability under an optimal sharing rule. Nota Di Lavoro, 77.2005. *Economics Energy Environment*. Fondazione Eni Enrico Mattei.

Weikard, H., Finus, M, & Altamirano-Cabrera, J.-C. (2006). The impact of surplus sharing on the stability of international climate agreements. *Oxford Economic Papers, 58*, 209–232.

Yi, S. S. (1997). Stable coalition structures with externalities. *Games and Economic Behavior, 20*, 201–237.