Finus, Michael; Rundshagen, Bianca

**Working Paper**

# How the Rules of Coalition Formation Affect Stability of International Environmental Agreements

Nota di Lavoro, No. 62.2003

**Provided in Cooperation with:**
Fondazione Eni Enrico Mattei (FEEM)

*Suggested Citation:* Finus, Michael; Rundshagen, Bianca (2003) : How the Rules of Coalition Formation Affect Stability of International Environmental Agreements, Nota di Lavoro, No. 62.2003, Fondazione Eni Enrico Mattei (FEEM), Milano

This Version is available at:
https://hdl.handle.net/10419/118097

# How the Rules of Coalition Formation Affect Stability of International Environmental Agreements

Michael Finus and Bianca Rundshagen

NOTA DI LAVORO 62.2003

**JULY 2003**

GG – Global Governance

Michael Finus, *Department of Economics, Hagen University, Germany*
Bianca Rundshagen, *Department of Economics, Hagen University, Germany*

# How the Rules of Coalition Formation Affect Stability of International Environmental Agreements


**Summary**

This paper compares stability of international environmental agreements for six different rules of coalition formation under very general conditions (any type of heterogeneity between countries). The rules can be interpreted as different institutional settings in which treaty formations take place and/or different designs of agreements. We consider open and restricted open membership game as well as four exclusive membership games with different degrees of unanimity required to form coalitions. From a policy perspective, counterintuitively, it turns out that stability is higher under exclusive than under open membership and stability increases with the degree of unanimity. We discuss the policy implications of our result for future treaty-making.



**Keywords:** Non-cooperative game theory, Rules of coalition formation, Stability

**JEL:** C72, H41, Q20



*Address for correspondence:*

Michael Finus
Department of Economics
University of Hagen
Profilstr. 8
58084 Hagen
Germany
E-Mail: Michael.Finus@fernuni-hagen.de, Bianca.Rundshagen@fernuni-hagen.de

## 1. Introduction

The game theoretical analysis of the formation and stability of international environmental agreements (IEAs) has become an important branch in the environmental economics literature over the last two decades. Roughly speaking, two approaches can be distinguished: repeated games and coalition games.[1] Repeated games analyze whether compliance with treaty obligations can be enforced in the long run with credible threats of punishment, invoking equilibrium concepts as for instance subgame perfect and renegotiation-proof equilibrium (Barrett 1994a, b, 1999, 2000, Endres/Finus 1998, Finus/Rundshagen 1998a, b, Finus/Tjotta 2002, Mäler 1994 and Stähler 1996). Coalition games analyze membership in IEAs, applying concepts of cooperative and non-cooperative game theory. The *cooperative approach* is based on the *characteristic function* that assigns a worth to coalitions. The worth is the aggregate payoff to a coalition that it can secure for itself irrespective of the behavior of countries outside the coalition. The focus of the analysis is on checking stability of the efficient grand coalition implementing a socially optimal emission vector, invoking the concept of the core (Chander/Tulkens 1995, 1997, Germain/Tulkens/de Zeeuw 1998 and Tulkens 1998). The efficient solution is stable (i.e., lies in the core) if no subgroup of countries has an incentive to form an other coalition, assuming that remaining countries break up into singletons playing either a minimax, maximin or Nash equilibrium strategy. In contrast, the *non-cooperative approach* is based on the *valuation function* that assigns an individual payoff to each country for each possible partition of countries, called coalition structure. For a fixed coalition structure payoffs follow from some assumption how countries choose their emissions. The standard assumption is that coalition members act as a single player maximizing the aggregate payoff to their coalition but behave non-cooperatively towards outsiders (see section 2). Equilibrium coalition structures are determined by applying the concept of internal and external stability (Barrett 1994b, 1997, Bauer 1992, Carraro/Siniscalco 1993, Hoel 1992, Hoel/Schneider 1997 Jeppesen/Andersen 1998 and Rubio/Ulph 2001). Internal stability means that no coalition member has an incentive to leave its coalition to become a singleton and external stability that no singleton has an incentive to join a coalition, assuming that the remaining countries do not revise their membership decision. Except Bauer (1992) all contributions have restricted coalition formation to a single coalition, allowing to group countries into signatories and non-signatories.

---

[1] For an overview of different approaches see Finus (2001 and 2003).

Recently, there has been a development of new approaches and concepts in game theory that we call "new coalition theory".[2] "New coalition theory" belongs to non-cooperative game theory and is based on the valuation function. Compared to the cooperative approach this has at least two advantages (see Bloch 1997). First, assuming that countries pursue their self-interest as rational players, it seems natural to conclude that countries will base their decision of membership on individual payoffs and not on the aggregate payoff to their coalition even if transfers are possible. Second, the valuation function does better account for spillovers between countries and coalitions. Since spillovers are an important source for free-riding in international treaty formation, the non-cooperative approach is better suited to rationalize inefficient IEAs, which, of course, most treaties are. Compared to the concept of internal and external stability these new developments have the advantage that they do not restrict ex-ante coalition formation to a single non-trivial coalition[3] but allow for the co-existence of multiple coalitions. Moreover, they invoke stability concepts that consider not only deviations by single countries but also by subgroups of countries where subgroups must not necessarily form one new coalition as assumed by the core but may form any partition. Finally, compared to the "classical" cooperative and non-cooperative approaches, new coalition theory draws a conceptual line between the rules of coalition formation (summarized in the definition of a coalition game) and stability (summarized in the definition of an equilibrium concept). Thus, differences in equilibrium coalitions can be unambiguously attributed to different stability concepts and different rules of coalition formation. Since the rules of coalition formation can be interpreted as the institutional setting in which treaty formation take place (Ecchia/Mariotti 1998), policy conclusions about the optimal design of agreements are possible. Moreover, the reaction of countries after a deviation do not follow from ad hoc assumptions but follow from the rules of coalition formation and can therefore be related to the rational behavior of countries.

This paper is in the tradition of new coalition theory and considers six different rules of coalition formation. Those rules allow for a comparison between open versus restricted open membership, open and restricted open versus exclusive membership and in the case of exclusive membership between different degrees of unanimity required to form coalitions. From a policy perspective counterintuitively, it turns out that stability is higher under

---

[2] For an excellent overview see Bloch (1997) and Yi (1997). Applications in the context of IEAs may be found in Carraro (2000), Carraro/Marchiori (2002), Carraro/Marchiori/Oreffice (2001), Finus (2002) and Finus/Rundshagen (2001).

[3] A non-trivial coalition means a coalition of at least two countries.

exclusive than under open membership and stability increases with the degree of unanimity. Moreover, we argue that restricted open membership is better suited than open membership to depict treaty-formation and has also theoretical advantages. In contrast to other papers, we are not interested in characterizing equilibria in economic and ecological terms (see the applied literature cited in footnote 2) since the purpose of this paper is to compare stability under very general conditions for single and multiple deviations. In what follows, we present our model in section 2, compare stability in section 3, and conclude with some remarks in section 4.

## 2. Model

### 2.1 Introduction

Coalition formation is modeled as a *two-stage game*. In the *first stage* countries decide on their *membership* in a coalition, in the *second stage* coalition members choose their *emissions*. The decision in the first stage depends on the *rules of coalition formation* that follow from the *definition of a coalition game*. The definition comprises three elements: 1) the set of players I={1, …, N} with a particular player denoted by index i or j, 2) the set of coalition or membership strategies $\Sigma = \Sigma_1 \times \Sigma_2 \times ... \times \Sigma_N$ with a particular strategy of country $i \in I$ denoted by $\sigma_i \in \Sigma_i$, and 3) a *coalition function* $\psi$ that maps membership strategies $\sigma = (\sigma_1, ..., \sigma_N)$ into coalition structures, $\psi : \Sigma \to C, \sigma \mapsto \psi(\sigma)$. A coalition structure $c = (c^1, ..., c^M)$ is a partition of players where a particular coalition is denoted by $c^k$, $k \in \{1, ..., M)$, $c^k \cap c^l = \varnothing$ $\forall$ k≠l, $\bigcup c^i = I$ and $c \in C$ where C is the set of all possible coalition structures. The decision in the second stage depends on the rules of emission choices that follow from the definition of the *valuation function*. The valuation function w maps coalition structures into a vector of individual payoffs (i.e., welfare) via an instruction how countries choose their emissions $e_i \in E_i$, $E = E_1 \times ... \times E_N$, for a given coalition structure c. Hence, the valuation function is a composition of two functions $w = \pi \circ \varepsilon$ where $\varepsilon : C \to E$, $c \mapsto \varepsilon(c)$ is a function mapping coalition structures into a vector of emission levels and $\pi : E \to \mathbb{R}^N$, $e \mapsto \pi(e)$ is a function mapping emission levels into welfare levels.

For the first stage we consider six different coalition games, representing six different institutional rules how coalitions form. For the second stage we consider only one rule that assigns a unique vector of individual payoffs for each possible coalition structure. This implies that countries choosing a membership strategy in the first stage know for each coalition game the implications in the second stage. Hence, games can be solved by backwards induction. Consequently, we start in the following by describing first stage 2 and subsequently we move on to explain stage 1 of the coalition formation game.

## 2.2 Second Stage of the Coalition Formation Game

Let the payoff to country i in the global emission game be given by

$$[1] \qquad \pi_i = \beta_i(e_i) - \phi_i(\sum_{j=1}^{N} e_j)$$

where we assume the following standard properties to hold (see, e.g., Folmer/van Mouche 2000): $\forall\ i \in I$ and $e_i \in [0, e_i^{max})$: $\beta_i' > 0$, $\beta_i'' < 0$, $\phi_i' > 0$, $\phi_i'' \geq 0$ where primes denote first and second derivative of a function. That is, benefits from emissions (in the form of consumption and production of goods), $\beta_i(e_i)$, increase at a decreasing rate. Damages from global emissions, $\phi_i(\sum_{i=1}^{N} e_j)$, increase in global emissions at a constant or increasing rate. Following the mainstream in the literature (e.g., Bloch 1997 and Yi 1997), we define the valuation function - mapping coalition structures into payoffs - as follows:

**Definition 1: Valuation Function**

*Fix a coalition structure* $c = (c^1, ..., c^M)$, *let* $w(c) = \{w_i(c^i; c)\}_{i \in I} = \{\pi_i(\varepsilon(c))\}_{i \in I} = \{\pi_i(e^*)\}_{i \in I}$ *and assume for instruction* $\varepsilon$ *that all players belonging to a coalition* $c^k \in c$ *jointly maximize the aggregate payoff to their coalition. Let* $e^k$ *denote the emission vector of coalition* $c^k$ *and* $e^{-k}$ *the emission vector of all other coalitions* $c^l \in c, l \neq k$, *and assume that equilibrium emission vector* $e^* = \varepsilon(c)$ *for coalition structure c satisfies:*

$\forall\ c^k \in C,\ \forall\ e^k \in E^k$: $\sum_{i \in c^k} \pi_i(e^*) \geq \sum_{i \in c^k} \pi_i(e^k, e^{-k*})$ *where* $e^*$ *is assumed to be a unique interior equilibrium.*

Definition 1 implies that the valuation of country i, $w_i(c^i; c)$ is identified by the entire coalition structure c and not only by the coalition to which country i belongs.[4] Countries behave cooperatively within their coalition but non-cooperatively against countries belonging to other coalitions when choosing their emissions. Put differently, members of a coalition act as one single player maximizing the aggregate welfare to their coalition and coalitions play a Nash equilibrium strategy in terms of emissions. Consequently, the singleton coalition structure (grand coalition) implies an equilibrium emission vector corresponding to the "classical" Nash equilibrium (social optimum). Uniqueness of the emission vector $e^*$ for each coalition structure $c \in C$ is related to the properties of the benefit and damage cost functions and follows from standard theorems. The assumption of an interior equilibrium eases

---

[4] Of course, it would be sufficient to write only $w_i(c)$, but it turns out that $w_i(c^i; c)$ is more convenient for later proofs.

establishing a fundamental feature of coalition formation in the context of IEAs that we call positive externality property and define and prove in Proposition 1.

In the case of transfers $\hat{w}_i(c^i;c) = w_i(c^i;c) + t_i$ where $t_i > 0$ implies to receive a transfer and $t_i < 0$ to pay a transfer. Typically, transfers follow from some assumption how coalitions share the gains from cooperation. For instance, in the context of the core the Chander/Tulkens´ transfer scheme (Chander/Tulkens 1995 and 1997) and in the context of internal&external stability the Shapley value (Botteon/Carraro 1997 and Barrett 1997) has been applied. Both transfer schemes assume that transfers are only exchanged between coalition members and that transfers balance, i.e., $\sum_{i \in c^i} t_i = 0$. For our purpose it suffices to show that there is a class of transfer schemes that preserves the positive externality property. For establishing this property, we need two lemmas.

**Lemma 1: Merging of Coalitions and Global Emissions**

*Let a coalition structure with M coalitions be denoted by* $c = (c^1,...,c^M)$*, a coalition structure with M-1 coalitions by* $c' = (c^1,...,c^{M-1})$ *where* $c'$ *is derived by merging two coalitions in c, and denote total emissions by* $e^{T*} = \sum_{i=1}^{N} e_i^*$*, then* $e^{T*}(c') < e^{T*}(c)$*.*[5]

**Proof:** Let $c^i$ and $c^j$ be two coalitions that merge and $c^k$ a coalition that is not involved in the merger. Assume $e^{T*}(c) < e^{T*}(c')$ instead of $e^{T*}(c) > e^{T*}(c')$ would be true after coalitions $c^i$ and $c^j$ have merged. Then, given the assumptions of the valuation function (Definition 1) and the assumptions about the properties of the benefit and damage cost functions, the following must be true:

$$\forall\ k \in c^k:\ \beta_k'(e_k^*(c)) = \sum_{k \in c^k} \phi_k'(e^{T*}(c)) \le \sum_{k \in c^k} \phi_k'(e^{T*}(c')) = \beta_k'(e_k^*(c')) \Rightarrow e_k^*(c) \ge e_k^*(c')$$

$$\forall\ i \in c^i \cup c^j:\ \beta_i'(e_i^*(c)) = \sum_{i \in c^i} \phi_i'(e^{T*}(c)) < \sum_{i \in c^i} \phi_i'(e^{T*}(c)) + \sum_{j \in c^j} \phi_j'(e^{T*}(c)) \le$$

$$\sum_{i \in c^i} \phi_i'(e^{T*}(c')) + \sum_{j \in c^j} \phi_i'(e^{T*}(c')) = \beta_i'(e_i^*(c')) \Rightarrow e_i^*(c) > e_i^*(c')$$

which obviously violates the initial assumption of $e^{T*}(c) < e^{T*}(c')$ (**Q.E.D.**).

Thus, Lemma 1 states that whenever singletons or non-trivial coalitions form a joint coalition, called a merger, global emissions will decrease. An immediate consequence of Lemma 1 is that the grand coalition implies the lowest and the singleton coalition structure the highest global emissions. Any coalition structure between these two benchmarks will imply lower

---

[5] For reference reason we mention that in the terminology of coalition theory coalition structure $c'$ is called coarser than coalition structure c.

global emissions than the singleton coalition structure but higher global emissions than the grand coalition. The next lemma looks at the reaction of outsiders to a merger.

**Lemma 2: Merging of Coalitions and Emissions of Outsiders**

*Let $c^k$ be a coalition that is not involved in a merger which implies that coalition structure c changes to coalition structure $c'$ and denote emissions of a member of coalition $c^k$ by $e_k$, then $e_k^*(c) \leq e_k^*(c')$.*

**Proof:** From Lemma 1 we have $e^{T*}(c) > e^{T*}(c')$ after a merger. Hence, the following must be true: $\forall\ k \in c^k : \beta_k'(e_k^*(c)) = \sum_{k \in c^k} \phi_k'(e^{T*}(c)) \leq \sum_{k \in c^k} \phi_k'(e^{T*}(c')) = \beta_k'(e_k^*(c')) \Rightarrow e_k(c) \leq e_k(c')$ (**Q.E.D.**).

Thus, Lemma 2 states that outsiders not involved in a merger will increase their emissions ($\phi_k'' > 0$) or at best do not revise their emission choices ($\phi_i'' = 0$). Lemma 1 and Lemma 2 allow stating the following result.

**Proposition 1a: Positive Externality Property (PEP): No Transfers**

*Assume no transfers and let $c^k$ be a coalition not involved in a merger which implies that coalition structure c changes to coalition structure $c'$, then in the global emission game $\forall k \in c^k : w_k(c^k; c) < w_k(c^k; c')$.*

**Proof:** Global emissions decrease after a merger by Lemma 1 and outsiders emission increase or remain unchanged by Lemma 2, hence benefits of outsiders increase or remain unchanged and their damage costs decrease. Consequently, outsiders´ welfare must increase through a merger (**Q.E.D.**).

The striking feature about Proposition 1a is that in the global emission game the PEP holds at a very general level, that is, for any type of welfare function and any type of heterogeneity between countries. The PEP has an immediate implication for stability of coalition structures in the various coalition games considered in the next subsection: if a country or group of countries change their membership strategy, that is, they leave their coalition, join an other coalition or form their own coalition, the harshest possible punishment is that all coalitions of the remaining countries break up into singletons. We will discuss this issue in more detail in section 3 and turn now to transfers.

For our level of generality it suffices to show that there exist a transfer scheme for which the PEP holds. We consider a modification of the transfer scheme proposed by Chander/Tulkens (1995 and 1997) which comes close to that applied in Eyckmans/Tulkens (1999):

$$[2] \qquad t_i = -\left[ w_i(c^i; c) - w_i(\{i\}; (1,...,1)) \right] + \lambda_i \left[ \sum_{i \in c^i} w_i(c^i; c) - \sum_{i \in c^i} w_i(\{i\}; (1,...,1)) \right]$$

where $\lambda_i$ are distributional weights, $0 < \lambda_i < 1$ and $\sum_{i \in c^i} \lambda_i = 1$, so that transfers balance within coalition $c^i$. The first term in brackets puts each country back to its payoff in the singleton coalition structure c=(1,...,1), corresponding to the classical Nash equilibrium. The second term gives each member of coalition $c^i$ a portion of the total gains (or losses) to this coalition from moving from the singleton coalition structure to coalition structure c. Losses cannot be generally ruled out since if some countries form a coalition, external countries may adjust their emissions upward according to Lemma 2. If this leakage effect is strong enough, cooperation may not be beneficial (for the entire coalition).[6] However, this does not affect the PEP as shown below.

**Proposition 1b: Positive Externality Property (PEP): Transfers**

*Assume transfer scheme [2] for all $i \in I$, let $\hat{w}_i(c^i; c) = w_i(c^i; c) + t_i$ and let $c^k$ be a coalition not involved in a merger, which implies that coalition structure c changes to coalition structure $c'$, then in the global emission game $\forall k \in c^k: \hat{w}_k(c^k; c) < \hat{w}_k(c^k; c')$.*

**Proof:** Computing $\hat{w}_k(c^k; c) = w_k(c^k; c) + t_k$ for [2] gives $\hat{w}_k(c^k; c) = w_k(\{k\}; (1,...,1)) + \lambda_k(\sum_{k \in c^k} w_k(c^k; c) - \sum_{k \in c^k} w_k(\{k\}; (1,...,1)))$. Since $w_k(\{k\}; (1,...,1))$ is unaffected by a merger and $w_k(c^k; c)$ increases through a merger by Proposition 1a, $\hat{w}_k(c^k; c)$ increases through an external merger (**Q.E.D.**)

From the proof it is evident that there are many transfer schemes that preserve the PEP. For instance, [2] measures the gains or losses compared to the singleton coalition structure. However, any other benchmark is also fine as long as it is a fixed benchmark. The weights $\lambda_i$ may be derived from any allocation rule as long as weights are not dependent on the partition of external players $I \setminus c^i$.[7]

## 2.3 First Stage of the Coalition Formation Games

In this subsection we define and discuss six coalition games that imply different rules how coalitions can form. All games assume that countries *simultaneously* announce their coalition

---

[6] Technically speaking, this implies that superadditivity may not hold in our context. Only for linear damage cost functions there are no leakages and hence superadditivity generally holds.

[7] In Chander/Tulkens (1995 and 1997) weights are related to marginal damage costs (i.e., $\lambda_i = \phi_i'(e^{T*}) / \sum_{i \in c^i} \phi_i'(e^{T*})$; notation of Lemma 1) and hence merging of coalitions generally affects weights. Thus, general conclusions about PEP are only possible for the special case of linear damage cost functions (and of course symmetric players).

strategy and allow for the *co-existence of several coalitions*. That is, coalition formation is not ex-ante restricted to a single coalition as this is assumed for the concepts of internal&external stability and the core that have been widely applied in the literature on IEAs.[8] However, games differ in the strategy set and most importantly in the coalition function that maps coalition strategies into coalition structures. From the discussion it will be apparent that there are two important features in which games differ. The first feature is membership where we distinguish between open and exclusive membership. Open membership means that countries can join any coalition they want whereas exclusive membership implies that this is only possible with the consent of the members of a coalition. The second feature is the degree of unanimity required to form a coalition. We distinguish four variants: weak, middle, strong and super strong.

The first game is called *open membership game* (OMG) and is due to Yi/Shin (1995). In this game countries can freely form coalitions as long as no outsider is excluded from joining a coalition. Countries choose their membership by announcing an address, i.e., a number between 1 and N. Countries that have announced the same address form a coalition. For instance, suppose N=4 and $\sigma_1 = \sigma_2 = \sigma_3 = 1$ and $\sigma_4 = 2$, then c={{1, 2, 3}, {4}} forms. If country 3 were to announce $\sigma_3 = 2$ instead, then c={{1, 2}, {3, 4}} would form. More formally, we have:[9]

**Definition 2: Open Membership Game (OMG)**

*a) The set of coalition strategies of country $i \in I$ is given by $\Sigma_i = \{1, ..., N\}$ where a particular strategy $\sigma_i$ is an announcement of an address.*

*b) Coalition function $\psi^{OMG}$ maps strategy vector $\sigma$ into coalition structure c as follows:*

$c^i = \{i\} \cup \{j \mid \sigma_i = \sigma_j\}$.

Thus in the OMG a country can join any coalition it wants. This strong assumption, however, seems not entirely in line with the notion of voluntary participation in IEAs that is one important feature reminiscent to the problem of cooperation in international pollution control. Hence, it seems natural to consider an extension of the OMG where countries have only unrestricted open access to non-trivial coalitions but require the consent of a single country if they intend to join it. This extension is called a *restricted open membership game* (ROMG). It

---

8 See the literature cited in the Introduction.

9 The rule of this game is similar to internal&external stability, except that in the open membership game multiple coalitions may form. For details see Finus/Rundshagen (2001).

has been proposed by Bloch (1997) and formalized by Rundshagen (2002). Conceptually, only a slight modification of Definition 2 is required, adding to the strategy set an address 0 and specifying the coalition function such that countries announcing 0 remain singletons.

**Definition 3: Restricted Open Membership Game (ROMG)**

*a) The set of coalition strategies set of country $i \in I$ is given by $\Sigma_i = \{0, 1, ..., N\}$ where a particular strategy $\sigma_i$ is an announcement of an address.*

*b) Coalition function $\psi^{ROMG}$ maps strategy vector $\sigma$ into coalition structure c as follows:*

$c^i = \{i\} \cup \{j \mid \sigma_i = \sigma_j \neq 0\}$.

For instance, recall our previous example that assumed $\sigma_1 = \sigma_2 = \sigma_3 = 1$ and $\sigma_4 = 2$ so that c={{1, 2, 3}, {4}} forms and where we argued that if country 3 changed its address to $\sigma_3 = 2$, then c={{1, 2}, {3, 4}} will come about. In the ROMG player 4 can announce $\sigma_4 = 0$ instead of $\sigma_4 = 2$ so that no other player can force him into a coalition. However, also in the ROMG, any player not in coalition {1, 2, 3} can join this coalition. This is different in the next four exclusive membership games.

In the *exclusive membership Δ-game* (EMΔG), which is due to Hart/Kurz (1983), countries announce a list of countries with which they like to form a coalition. Those countries that announce the same list will form a coalition. For instance, suppose N=4 and $\sigma_1 = \{1, 2, 3\}$, $\sigma_2 = \{1, 2, 3\}$, $\sigma_3 = \{3\}$ and $\sigma_4 = \{1, 2, 3, 4\}$, then $c = \{\{1, 2\}, \{3\}, \{4\}\}$ forms. Countries 1 and 2 propose the same list and therefore form a coalition. Countries 3 and 4 remain singletons. Though country 4 would like to form a coalition with all other countries, country 3 can remain a singleton and country 1 and 2 form their own coalition since *membership is exclusive*. In other words, a coalition only forms by *unanimous agreement*. More formally, we have:

**Definition 4: Exclusive Membership Δ-Game (EMΔG)**

*a) The set of coalition strategies of country $i \in I$ is given by $\Sigma_i = \{c^i \subset I \mid i \in c^i\}$ where a particular strategy $\sigma_i$ is a list of countries with which country i would like to form a coalition.*

*b) Coalition function $\psi^{EM\Delta G}$ maps strategy vector $\sigma$ into coalition structure c as follows:*

$c^i = \{i\} \cup \{j \mid \sigma_i = \sigma_j\}$.

As it will turn out from a comparison with other exclusive membership games below, in the EMΔG only a *weak degree of unanimity* is required to form a coalition.

The *exclusive membership Γ-game* (EMΓG) goes back to Von Neumann/Morgenstern (1944) and has been reintroduced by Hart/Kurz (1983) under this name. This game is identical to the last game in terms of strategies but different in terms of the coalition function. Whereas in the EMΔG it suffices that a subgroup of countries on a list makes the same proposal (and hence the subgroup forms a coalition), in the EMΓG a coalition forms *if and only if* all members on a list make the same proposal. That is, the degree of unanimity to form a coalition in the EMΓG is higher than in the EMΔG. For instance, suppose our previous example in the context of the EMΔG that assumed $\sigma_1 = \{1, 2, 3\}$, $\sigma_2 = \{1, 2, 3\}$, $\sigma_3 = \{3\}$ and $\sigma_4 = \{1, 2, 3, 4\}$, which led to coalition structure $c = \{\{1, 2\}, \{3\}, \{4\}\}$, whereas now it implies $c = \{\{1\}, \{2\}, \{3\}, \{4\}\}$. If and only if country 1 and 2 were to announce $\sigma_1 = \sigma_2 = \{1, 2\}$ would they form a coalition. More formally, we define:

**Definition 5: Exclusive Membership Γ-Game (EMΓG)**

*a) The set of coalition strategies of country $i \in I$ is given by $\Sigma_i = \{c^i \subset I \mid i \in c^i\}$ where a particular strategy $\sigma_i$ is a list of countries with which country i would like to form a coalition.*

*b) Coalition function $\psi^{EM\Gamma G}$ maps strategy vector $\sigma$ into coalition structure c as follows:*

*$c^i = \sigma_i$ if and only if $\sigma_i = \sigma_j \ \forall j \in \sigma_i$, otherwise $c^i = \{i\}$.*

In comparison to the subsequent games we call this a *middle degree of unanimity* to form a coalition.

The *exclusive membership H-game* (EMHG) has been invented by Finus/Rundshagen (2003) and implies a modification not only in terms of the coalition function but also in terms of strategies compared to the EMΔG and EMΓG. In terms of coalition strategies, countries´ announcements comprise not only a list of countries with which they would like to form a coalition but also a *list with their preferred residual coalition structure*. The coalition function determines coalition structure c not in one but in two steps. The first step resembles that in the EMΓG: countries which have each other on the list to form a coalition will be in one coalition if and only if all members on a list make the same proposal. This leads to a "preliminary" coalition structure $\tilde{c}$. The second step requires that all members belonging to coalition $\tilde{c}^i$ in $\tilde{c}$ have correctly announced the external coalitions $\tilde{c}^j, ...., \tilde{c}^k$ forming in $\tilde{c}$ otherwise $\tilde{c}^i$ splits up into singletons in the "final" coalition structure c. For instance, suppose N=5 and the following announcements: $\sigma_1 = \sigma_2 = \{\{1,2\}, \{3\}, \{4\}, \{5\}\}$, $\sigma_3 = \sigma_4 = \{\{1,2\}, \{3,4\}, \{5\}\}$ and $\sigma_5 = \{\{1,2\}, \{3,4,5\}\}$. Thus, in the first step, preliminary coalition structure $\tilde{c} = \{\{1,2\}, \{3,4\}, \{5\}\}$ forms since countries 1 and 2 and 3 and 4 propose exactly the same list

with which countries they would like to form a coalition. In the second step final coalition structure $c = \{\{1\},\{2\},\{3,4\},\{5\}\}$ follows since only countries 3 and 4 announcement materializes in $\tilde{c}$. More specifically:[10]

**Definition 6: Exclusive Membership H-Game (EMHG)**

*a) The set of coalition strategies of country $i \in I$ is given by $\Sigma_i = \{c(i) \in C / i \in c^1(i)\}$ where a particular strategy $\sigma_i = c(i) = (c^1(i); c^2(i), ..., c^{M_i}(i))$ of country i is composed of a list of countries with which it wants to form a coalition, $c^1(i)$, and its preferred residual coalition structure, $c^2(i), ..., c^{M_i}(i)$.*

*b) Coalition function $\psi^{EMHG}$ determines coalition structure c from strategy vector $\sigma$ in two steps as follows.*

*First, a preliminary coalition structure $\tilde{c} = (\tilde{c}^1, ..., \tilde{c}^{\tilde{M}})$ is determined: $i \in \tilde{c}^1$ if only if $c^1(i) = c^1(j)$ $\forall j \in c^1(i)$, otherwise $\tilde{c}^1 = \{i\}$.*

*Second, final coalition structure $c = (c^1, c^2, ..., c^M)$ follows from: $\tilde{c}^j \in c \Leftrightarrow c(i) = \tilde{c}$ $\forall i \in \tilde{c}^j$, otherwise $\tilde{c}^j$ splits up into singletons in c.*

Thus, the whole coalition formation process can be interpreted as follows. In the first step it is checked whether *"internal lists"* match, that is, lists of countries with which a country wants to form a coalition. The preliminary formation process requires a degree of unanimity of the Γ-type. In the second step it is checked whether *"external lists"* match, that is, lists of partitions formed by external countries. Here, only lists of members in the same coalition but not of all countries must match to form a coalition. This implies de facto that a degree of unanimity of the Δ-type with respect to the external list is required to form a coalition. This suggests that a game can be constructed which also requires a degree of unanimity of the Γ-type for the external list. This is done below. For reference reason we call the *degree of unanimity* required to form a coalition in terms of the entire EMHG *"strong"* in order to distinguish it from that in the next game that we call *"super strong"*.

The *exclusive membership I-game* (EMIG) is due to Finus/Rundshagen (2003) and is identical to the EMHG in terms of strategies but different in terms of the coalition function. In this game not only must all members of a coalition propose the same external list but all countries to form coalitions. Thus, taken together, a coalition only forms if and only if all countries make the same proposal for the entire coalition structure, comprising an internal and external

---

[10] There is a close resemblance between core-stability and a strong Nash equilibrium in this game. For details see Finus/Rundshagen (2003).

list. For instance, reconsider the example in the context of the EMHG that assumed N=5 and announcements $\sigma_1 = \sigma_2 = \{\{1,2\},\{3\},\{4\},\{5\}\}$, $\sigma_3 = \sigma_4 = \{\{1,2\},\{3,4\},\{5\}\}$ and $\sigma_5 = \{\{1,2\},\{3,4,5\}\}$ and where in the H-game $c = \{\{1\},\{2\},\{3,4\},\{5\}\}$ formed. In contrast, in the EMIG $c = \{\{1\},\{2\},\{3\},\{4\},\{5\}\}$ because not all announcements with respect to the external list match. Formally, the coalition function determines coalition structures in one step:

**Definition 7: Exclusive Membership I-Game (EMIG)**

*a) The set of coalition strategies of country $i \in I$ is given by $\Sigma_i = \{c(i) \in C / i \in c^1(i)\}$ where a particular strategy $\sigma_i = c(i) = (c^1(i); c^2(i), ..., c^{M_i}(i))$ of country i is composed of a list of countries with which it wants to form a coalition, $c^1(i)$, and its preferred residual coalition structure, $c^2(i), ..., c^{M_i}(i)$.*

*b) Coalition function $\psi^{EMIG}$ determines coalition structure c from strategy vector $\sigma$ as follows:*

*$c = c(i)$ if and only if $\sigma_i = \sigma_j \ \forall i \in I$, otherwise $c = (1,...,1)$.*

After the discussion of how coalitions form in the various coalition games, we can now turn to analyze stability.

## 3. Stability of Coalition Structures

### 3.1 Introduction

In this section we compare stability of coalition structures in the six coalition games. We consider two equilibrium concepts: Nash equilibrium and strong Nash equilibrium. A Nash equilibrium coalition structure, abbreviated NE, is derived from a vector of coalition strategies $\sigma^*$ where no single country has an incentive to change its strategy (announcement), given that other countries announce their equilibrium strategy. This is the familiar definition of Nash equilibrium, except that strategies are coalition and not economic (i.e., emission) strategies. Similar, a strong Nash equilibrium coalition structure, abbreviated SNE, is a vector of coalition strategies where no subgroup of countries $I^S \subset I$ has an incentive to change its coalition strategy. Formally, we have:

**Definition 8: Nash and Strong Nash Equilibrium Coalition Structures**[11]

*Let $\hat{C}^{I^S}(\sigma)$ be the set of coalition structures that a subgroup of countries $I^S$ can induce if the remaining countries $j \in I \backslash I^S$ play $\sigma^{I \backslash I^S}$. Then $\sigma^*$, inducing coalition structure $c^*$, is called a SNE if no subgroup $I^S$ can increase its members' payoff by inducing another coalition structure $\hat{c} \in \hat{C}^{I^S}(\sigma^*)$. That is, $c^*(\sigma^*)$ is a SNE if there is no $I^S \subset I$ and a coalition structure $\hat{c} \in \hat{C}^{I^S}(\sigma^*)$ such that $w_i(\hat{c}^i, \hat{c}) \geq w_i(c^i, c^*) \ \forall i \in I^S$ and $\exists j \in I^S$: $w_j(\hat{c}^j, \hat{c}) > w_j(c^j, c^*)$. For a NE, $I^S = \{i\}$.*

Given that multiple deviations are a special case of single deviations ($I^S=\{i\}$), it is evident that the set of SNE is a subset of NE, $C^{SNE} \subset C^{NE}$. The reason why we consider not only SNE but also NE is that existence of NE is guaranteed under far more general conditions than of SNE. Since some of the proofs in subsection 3.2 are instructive for establishing existence of equilibrium coalition structures in the various coalition games, we postpone the discussion until subsection 3.3.

### 3.2 Comparison of Equilibrium Coalition Structures

For the analysis of stability it is helpful to note four things in advance. First, stability is defined in terms of incentives to induce other coalition structures. Possible inducements follow from the rules of coalition formation and may be broken down in two components. a) The deviations that are available to a country or group of countries if they change their coalition strategies. This direct effect comprises the possibility of deviators forming new coalitions and/or joining other coalitions. b) The reaction to a deviation of those countries not (actively) involved in a deviation. This indirect effect comprises reactions ranging from no reaction to the resolution of all partitions to which non-deviating countries belong. Second, trivially, when checking stability of coalition structure c, we are only interested in changes of strategies that will have an effect on c. Third, when comparing SNE in the various games, we only have to consider deviations by a true subgroup of countries $I^S \subsetneq I$ since deviations by all countries can induce any coalition structure in every game. Thus, if there are differences in stability in the various games, they must stem from the possibilities that are available to subgroups of countries. Fourth, several coalition strategy vectors may lead to the same

---

[11] We define strong Nash equilibrium in terms of a weak inequality to be consistent with the definition of Pareto-optimal coalition structures in subsection 3.3. A modification of this assumption would not affect the subsequent proofs.

coalition structure.[12] Hence, when analyzing stability of a coalition structure, it suffices that stability holds for one coalition strategy vector. Since the most favorable condition for stability in the exclusive membership Δ- and Γ-game is if each coalition member proposes exactly this coalition to which it belongs in c and in the exclusive membership H- and I-game if each country announces exactly coalition structure c, our proofs start from this assumption.

In the following we proceed in three steps to derive our final result. First, we compare equilibrium coalition structures in the open membership game (OMG), restricted open membership game (ROMG) and in the exclusive membership Δ-game (EMΔG) since these three games differ only in the direct but not in the indirect effect of deviations. Second, we compare equilibrium coalition structures in the exclusive membership Δ-, Γ- and H-game (EMΔG, EMΓG and EMHG) since these games do not differ in the direct but in the indirect effect. Third, we contrast equilibrium coalition structures in the exclusive membership H-game with those in exclusive membership I-game (EMIG) since these games differ only in the direct effect. We immediately start with the first comparison.

**Proposition 2: Comparison of Equilibria in the OMG, ROMG and EMΔG**

*Let the set of Nash and strong Nash equilibrium coalition structures (NE and SNE) in the OMG, ROMG and EMΔG be denoted by* $C^{NE}(...)$ *and* $C^{SNE}(...)$ *respectively, then*

*a)* $C^{NE}(OMG) \subset C^{NE}(ROMG) \subset C^{NE}(EM\Delta G)$ *and*

*b)* $C^{SNE}(OMG) \subset C^{SNE}(ROMG) \subset C^{SNE}(EM\Delta G)$.

**Proof:** First we show that if $c \notin C^{SNE}(ROMG)$, then $c \notin C^{SNE}(OMG)$. Suppose $c = (c^1,...,c^M) \notin C^{SNE}(ROMG)$, then there exists a group of countries $I^S \subset I$ and a set of announcements $\sigma^{I^{S'}}$ such that $w_i(c^{i'};c') \geq w_i(c^i;c)\ \forall\ i \in I^S$ and $\exists\ j \in I^S : w_j(c^{j'};c') > w_j(c^j;c)$ holds where $c' = c'(\sigma^{I^{S'}}, \sigma^{I \setminus I^{S*}})$. Since $|\Sigma| = N$ there exists $\sigma^{I^{S''}} \neq 0\ \forall i \in I^S$ in the OMG leading to $c'' = c''(\sigma^{I^{S''}}, \sigma^{I \setminus I^{S*}})$ so that $c' = c''$. Hence $c'$ can also be induced in the OMG and $c \notin C^{SNE}(OMG)$ follows. For a NE, the same reasoning applies with $I^S = \{i\}$.

Second we show that if $c \notin C^{SNE}(EM\Delta G)$, then $c \notin C^{SNE}(ROMG)$. Suppose $c = (c^1,...,c^M) \notin C^{SNE}(EM\Delta G)$, then there exists a group of countries $I^S \subset I$ for which $w_i(c^{i'};c') \geq w_i(c^i;c)\ \forall\ i \in I^S$ and $\exists\ j \in I^S : w_j(c^{j'};c') > w_j(c^j;c)$ holds where $c' = c^S \cup (c^1 \cap (I \setminus I^S)) \cup ... \cup (c^M \cap (I \setminus I^S))$ and $c^S = (c^{S_1},...,c^{S_L})$ is a partition of $I^S$ with $c^{S_i}$ a

[12] For instance, in the open membership game announcements $\sigma_1 = \sigma_2 = 1$ and $\sigma_2 = 2$ lead to the same coalition structure c=({1,2}, {3}) as announcements $\sigma_1 = \sigma_2 = 2$ and $\sigma_2 = 3$. Note, however, that in each coalition game each strategy vector leads to a unique coalition structure.

particular coalition in $c^S$. Since $c'$ can also be induced in the ROMG (by countries $i \in c^{S_i}$ changing their announcements to $\sigma^{S_i} \neq 0$ and $\sigma^{S_i} \neq \sigma^j \ \forall\ j \notin c^{S_i}$), $c \notin C^{SNE}(\text{ROMG})$ follows. For a NE, the same reasoning applies with $I^S = \{i\}$. (**Q.E.D.**).

The intuition of Proposition 2 is the following. In all three games the indirect effect of a deviation is the same. If in the OMG and ROMG deviating countries change their address all other countries remain in their coalitions. In the EMΔG the deviating countries change their list of countries with which they like to form a coalition. Due to weak unanimity to form coalitions, other countries will remain in their coalitions if some countries change their list. However, the three games differ in their direct effect. The direct effect comprises that deviators form their own partition or a partition with other countries not actively involved in a deviation. The latter effect implies to join other coalitions. In the OMG any possible deviation is available to a subgroup of countries, in the ROMG deviators cannot join singletons without their consent and in the EMΔG deviators can neither join singletons nor non-trivial coalitions without their consent. Thus, it is easier to sustain a NE or SNE in the EMΔG than in the ROMG since the amount of possible deviations is smaller in the former than in the latter game, anything else being equal. The same is true when comparing ROMG and OMG. We now turn to the second comparison.

**Proposition 3: Comparison of Equilibria in the EMΔG, EMΓG and EMHG**

*Let the set of Nash and strong Nash equilibrium coalition structures (NE and SNE) in the EMΔG, EMΓG and EMHG be denoted by $C^{NE}(...)$ and $C^{SNE}(...)$, respectively, then*

*a)* $C^{NE}(EM\Delta G) \subset C^{NE}(EM\Gamma G) \subset C^{NE}(EMHG)$ *and*

*b)* $C^{SNE}(EM\Delta G) \subset C^{SNE}(EM\Gamma G) \subset C^{SNE}(EMHG)$.

**Proof:** Suppose in the EMΔG and EMΓG that each country announces exactly those countries with which it forms a coalition in c and in the EMHG each country announces exactly coalition structure c. Consider a deviation (change of strategies) by a subgroup of countries $I^S \subsetneq I$. Let $c = (c^T, c^R)$ be the initial coalition structure and $c' = (c^{T(1)'}, c^{T(2)'}, c^{R'})$ the resulting coalition structure after deviation. $c^{T(1)'}$ is the partition of deviators $I^S \subset I^T \subsetneq I$ in $c'$ that belonged to partition $c^T$ in c, $c^{T(2)'}$ is the partition of remaining countries $I^T \backslash I^S$ in $c'$ that belonged to partition $c^T$ in c, and $c^R$ is the partition of all remaining countries $I \backslash I^T$ before and $c^{R'}$ after the deviation. In the EMΔG and EMΓG a deviation has no effect on $c^R$ and hence $c^R = c^{R'}$. In the EMΔG a deviation implies that those coalitions to which the deviators belong stick together whereas in the EMΓG they break up into singletons by the stronger degree of unanimity required forming a coalition. Thus, if we let $c^T = (c^1, ..., c^L)$, then in the EMΔG

$c^{T(2)'} = c^1 \cap (I^T \setminus I^S) \cup ... \cup c^L \cap (I^T \setminus I^S)$ whereas in the EMΓG $c^{T(2)'} = (1,...,1)$. In the EMHG a deviation by a subgroup of countries $I^S \subset I^T$ leads to intended partition $\tilde{c}^{T(1)'}$ in the first step of the coalition function if $\forall i \in I^S \; \forall j \in c^k(i) : c^k(i) = c^k(j)$. However, all coalitions to which the deviators belonged will break up into singletons in $\tilde{c}'$ because their announcements do not match anymore. Thus, in the first step of the coalition function $\tilde{c}^{T(2)'} = (1,...,1)$. In the second step, all countries belonging to other coalitions (partition $c^R = \tilde{c}^{R'}$) will break up into singletons since those countries must have initially (before the deviation) announced the correct coalition structure that formed in $\tilde{c}$ otherwise they would not have been in coalitions in c. Now their announcements do not match anymore and hence those coalitions in partition $c^R = \tilde{c}^{R'}$ break apart in the second step of the coalition function. Hence, $c' = (c^{T(1)'}, c^{T(2)'}, c^{R'})$ with $c^{T(2)'} = (1,...,1)$ and $c^{R'} = (1,....,1)$.[13] Taken together, we have for a deviator $i \in c^{T(1)_i'}$, $c^{T(1)_i'} \in c^{T(1)'}$: $w_i(c^{T(1)_i'}; c^{T(1)'}, c^{T(2)'}, c^{R'}) \geq w_i(c^{i'}; c^{T(1)'}, c^{T(2)'} = (1,...,1), c^{R'}) \geq w_i(c^{i'}; c^{T(1)'}, c^{T(2)'} = (1,...,1), \; c^{R'} = (1,....,1))$ after the deviation due to the positive externality property. Hence, $C^{SNE}(EM\Delta G) \subset C^{SNE}(EM\Gamma G) \subset C^{SNE}(EMHG)$. For NE, the same reasoning applies with $I^S = \{i\}$ (**Q.E.D.**).

In all three games the direct effect of deviations is the same: deviators cannot join other coalitions due to exclusivity and the partition that a group of countries can form is the same. However, the indirect effect is different since the various degrees of unanimity required to form a coalition imply different reactions of the remaining players. In the EMΔG there is no reaction, in the EMΓG coalitions to which the deviators belonged break apart, and in the EMHG, additionally, all other coalitions break apart. Thus, the higher the degree of unanimity necessary to form coalitions, the higher is the implicit punishment in positive externality games after a deviation and hence the higher is the "degree of stability". We turn now to the last comparison.

**Proposition 4: Comparison of Equilibria in the EMHG and EMIG**

*Let the set of Nash and strong Nash equilibrium coalition structures (NE and SNE) in the EMHG and EMIG be denoted by $C^{NE}(...)$ and $C^{SNE}(...)$, respectively, then*

*a) $C^{NE}(EMHG) = C^{NE}(EMIG)$ and b) $C^{SNE}(EMHG) \subset C^{SNE}(EMIG)$.*

---

[13] Of course, partition $c^{T(1)'}$ will only form if all deviators $I^S$ correctly announce partition $\tilde{c}^{R'}$ and $\tilde{c}^{T(2)'}$ in $\tilde{c}'$. If not, then $c^{T(1)'}$ would break up into singletons, leading to $c' = (1,...,1)$, which can also be induced by a single deviation.

**Proof:** According to the proof of Proposition 3, in the EMHG multiple deviations from coalition structure c lead to $c' = (c^{T(1)'}, 1, ..., 1)$ which in terms of a single deviation implies $c^{T(1)'} = \{i\}$ and thus $c' = (1, ..., 1)$. In the EMIG any deviation by a subgroup of countries $I^S \subsetneq I$ leads to the complete resolution of all coalition structures including the partition of deviating countries since coalition strategies do not match anymore and hence $c' = (1, ..., 1)$. Thus, in terms of single deviation there is no difference between both games but in terms of multiple deviations: any partition that can be induced by $I^S$ in the EMIG can also be induced in the EMHG but not vice versa (**Q.E.D.**).

Summarizing Proposition 2, 3 and 4 gives our central result:

**Proposition 5: Comparison of Equilibria in All Coalition Formation Games**

*Let $C^{NE}(...)$ and $C^{SNE}(...)$ denote the set of Nash equilibrium (NE) and strong Nash equilibrium (SNE) coalition structures in the open membership game (OMG), the restricted open membership game (ROMG), and the exclusive membership Δ-, Γ-, H, and I-game (EMΔG, EMΓG, EMHG and EMIG), respectively, then*

*a)* $C^{NE}(OMG) \subset C^{NE}(ROMG) \subset C^{NE}(EM\Delta G) \subset C^{NE}(EM\Gamma G) \subset C^{NE}(EMHG) = C^{NE}(EMIG)$ *and*

*b)* $C^{SNE}(OMG) \subset C^{SNE}(ROMG) \subset C^{SNE}(EM\Delta G) \subset C^{SNE}(EM\Gamma G) \subset C^{SNE}(EMHG) \subset C^{NE}(EMIG)$.

**Proof:** Follows from Proposition 2, 3 and 4 (**Q.E.D.**).

Proposition 5 stresses the relation between the rules of coalition formation and stability of agreements with an unambiguous relation (inclusion chain) between equilibria in the various coalition games.

### 3.3 Existence of Equilibrium Coalition Structures

The reason for considering not only strong Nash equilibrium (SNE) but also Nash equilibrium (NE) coalition structures in the previous subsection is that - at our level of generality - existence of a NE is guaranteed in all games except in the open membership game whereas existence of a SNE can only be established in the exclusive membership I-game. We start establishing existence of a NE for which we need the following definition.

**Definition 9: Individually Rational Coalition Structures**

*A coalition structure c is called individually rational if each player receives at least his payoff in the singleton coalition structure, i.e.,* $\forall i \in I: w_i(c^i, c) \geq w_i(\{i\}, 1, ..., 1)$.

It is evident that the set of individually rational coalition structures, henceforth abbreviated $C^{IR}$, is non-empty since the singleton coalition structure belongs to this set by definition. Moreover, intuition suggests that there is a close relation between individually rational and Nash equilibrium coalition structures.

**Lemma 3: Individually Rational and Nash equilibrium Coalition Structures**

*In every coalition game a Nash equilibrium coalition structure must be individually rational, i.e.,* $C^{NE} \subset C^{IR}$.

**Proof:** Consider a coalition structure $c = (c^1, ..., c^M)$ and suppose that country i is a member of coalition $c^1$. In each coalition game country i can induce a coalition structure of type $c' = (\{i\}, c^{R'})$ by changing its strategy (where $c^R = c \setminus \{i\}$). In the worst case $c^R = (1, ..., 1)$ because of the positive externality property (PEP). Hence, a coalition structure c can only be a NE if and only if $\forall i \in I: w_i(c^i; c) \geq w_i(\{i\}; c')$ where $c' = (1, ...., 1)$ (**Q.E.D.**).

Using Lemma 3 and recalling the fact that the singleton coalition structure is individually rational by definition, it is evident that existence of a NE is guaranteed in the restricted open membership (ROMG) and the four exclusive membership games. If each country announces address $\sigma_i = 0$ in the ROMG, then no country can unilaterally induce any other coalition structure. The same is true if each country announces a list with only itself in the exclusive membership Δ- and Γ-game and if each country announces the singleton coalition structure in the exclusive membership H- and I-game. Hence, we can state the following proposition without proof.

**Proposition 6: Existence of Nash Equilibrium Coalition Structures**

*In the restricted open membership game and the four exclusive membership games a Nash equilibrium coalition structure exists.*

However, in the open membership game a NE may not always exist. Suppose each country announces a different address, then an individual country i that has an incentive to join an other singleton j can deviate by announcing the same address, $\sigma_i = \sigma_j$. The resulting coalition structure is also unstable if country j prefers to stay alone. In any non-trivial coalition structure stability may also be a problem since basically any deviation is possible in the open

membership game.[14] This stresses that the restricted open membership does not only capture voluntary participation better but has also theoretical advantages compared to the open membership game. We now turn to SNE for which we need the following definition.

**Definition 10: Pareto-optimal Coalition Structures**

*A coalition structure c is Pareto-optimal if there is no other coalition structure $c'$ where at least one country is better off and no country worse off, i.e., there is no $c'$ such that* $w_i(c^{i'}, c') \geq w_i(c^i, c) \;\; \forall i \in I \wedge \exists j \in I: \; w_j(c^{j'}, c') > w_j(c^j, c)$.

Definition 10 is the familiar definition of Pareto-optima, applied to the context of coalition formation. Henceforth, we abbreviate Pareto-optimal coalition structures by PO and denote its set by $C^{PO}$. It is evident that the grand coalition is always a PO: it generates the highest global welfare and therefore in any other coalition structure at least one country must be worse off. Consequently, $C^{PO}$ is non-empty. Definition 10 suggests that there is a close relation between PO and SNE: a deviation by subgroup of countries includes the special case of a deviation of all countries, $I^S$=I. Hence, a necessary condition that a coalition structure is a SNE is that it is a PO, $C^{SNE} \subset C^{PO}$. Moreover, recalling that $C^{SNE} \subset C^{NE}$ because multiple deviations include the special case of single deviations, $I^S$={i}, and that $C^{NE} \subset C^{IR}$ from Lemma 2, it is apparent that we can state the following lemma (without proof).

**Lemma 4: Pareto-optimal and Strong Nash Equilibrium Coalition Structures**

*In every coalition game a strong Nash equilibrium coalition structure must be an individually rational and Pareto-optimal coalition structure, i.e.,* $C^{SNE} \subset C^{IR} \cap C^{PO}$.

From Lemma 4 we see that existence of a SNE faces two problems. First, not any PO is individually rational. For instance, as it is well known, in the absence of compensation payments the socially optimal solution, which corresponds to the grand coalition in our context, may not be individually rational if countries have heterogeneous payoff functions. Of course, this problem can be mitigated by a transfer scheme as for instance the one we proposed in subsection 2.2, which ensures that at least the grand coalition is individually rational (apart from the singleton coalition structure).[15] Second, not any PO is a SNE. Despite

---

[14] An example of non-existence of a NE in the open membership game is provided in Yi/Shin (2000) in a theoretical context and in Eyckmans/Finus (2003) in an empirical context. In Yi/Shin (2000) conditions for existence are derived which are, however, very restrictive.

[15] Since the aggregate payoff in the grand coalition is higher than in any other coalition structure, individual rationality is ensured if each country receives a fraction of the gains from cooperation. However, in any other coalition structure it cannot be ruled out (except for linear damage cost

the fact that not all countries can benefit when moving from a PO to some other coalition structure, this may well be the case for a true subgroup $I^S \subsetneq I$. This is the reason why in many economic applications as well as in almost all coalition games analyzed in this paper a SNE may fail to exist. The only exception is the exclusive membership I-game. In the case of transfer scheme [2], this is easy to see. First, the grand coalition is individually rational and a PO. Second, any deviation by subgroup of countries $I^S \subsetneq I$ leads to $c' = (1,...,1)$ and is therefore not beneficial. However, existence holds at a far more general level.

**Proposition 7: Existence of SNE in the Exclusive Membership I-game**

*There always exists a strong Nash equilibrium in the exclusive membership I-game.*

**Proof:** We proceed in two steps. First, we show that the set of individually rational Pareto-optimal coalition structures is none empty, $C^{IR} \cap C^{PO} \neq \varnothing$. Second, we demonstrate that $C^{SNE}(EMIG) = C^{IR} \cap C^{PO}$. 1) Suppose that $c = (1,...,1) \in C^{PO}$ and recall that $c = (1,...,1) \in C^{IR}$. Then, $C^{IR} \cap C^{PO} \neq \varnothing$ is obvious. Alternatively, suppose $c = (1,...,1) \notin C^{PO}$. Then, there exists a coalition structure $c'$ that Pareto-dominates c. If $c' \in C^{PO}$, we are done, if not, then there exists a coalition structure $c''$ that Pareto-dominates $c'$. This process continues until a coalition structure is an element of $C^{PO}$ (otherwise assumption $c = (1,...,1) \notin C^{PO}$ must be wrong). 2) Any deviation by a subgroup of countries $I^S \subsetneq I$ leads to the singleton coalition structure, which is not beneficial if a coalition structure is individually rational. Any deviation by all countries $I^S$=I is not profitable if a coalition structure is Pareto-optimal (**Q.E.D.**).

Thus, we know that at least one element in the inclusion chain of SNE in Proposition 5 is none empty.[16]

## 4. Summary and Final Remarks

We analyzed coalition formation in the tradition of "new coalition theory" that has several advantages to former approaches. 1) The analysis is based on individual and not on aggregate payoffs of players. 2) Externalities among players and coalitions are fully captured. 3) A conceptual distinction between the rules of coalition formation and equilibrium is possible.

---

functions) that the aggregate payoff to a coalition is lower than the sum of coalition members´ payoffs in the singleton coalition structure and hence individual rationality may fail to hold. See the discussion in subsection 2.2.

[16] Existence of SNE in the exclusive membership Γ- and H-game can be established for the restrictive assumption of symmetric players as shown in Finus/Rundshagen (2001). However, even this restrictive assumption may not guarantee existence of SNE in the open membership games and in the exclusive membership Δ-game.

4) Coalition formation is not restricted ex-ante to a single (non-trivial) coalition and hence the co-existence of multiple coalition is possible. 5) Stability can be defined in terms of multiple deviations where deviators may form any partition they want.
We considered six coalition games that can be interpreted as different institutional settings in which coalition formation takes place and/or different designs of treaty protocols. We compared stability in the six games under very general conditions, applying the concept of a Nash equilibrium and strong Nash equilibrium. We demonstrated that this is possible based on only one condition called positive externality property that holds in the global emission game without transfers but also for a large class of transfer schemes.

Though our results are derived from a stylized model and despite results may seem obvious when considering the proofs, they are interesting from a policy perspective. They suggest that exclusive membership may be conducive to stability of IEAs. Given that almost all protocols of existing IEAs allow non-signatories to join an IEA at any time they want suggests altering this rule from open to exclusive membership for future IEAs. This basically would imply to turn a public good agreement into a club good agreement in terms of membership. Our results also suggest that a high degree of unanimity in terms of membership helps to stabilize an IEA whereas it is usually conjectured that unanimity leads to agreements of the lowest denominator type. The driving force in our model is that the higher the degree of unanimity required to form a coalition, the higher is the pressure on countries to accede to an agreement since a failure has severe consequences. Though in our model unanimity applies to membership and not to the level of emission reductions, our model indicates that the widespread application of unanimous decision rules in international politics may not always be a disadvantage and, in fact, may be a rational choice. In any case, our conclusion is in line with bargaining models that have analyzed the positive effect of unanimous decision rules in terms of the level of emission reduction and the policy instrument used to implement emission reduction targets (Endres 1997 and Finus/Rundshagen 1998a, b).

Of course from an economic and ecological perspective it would be interesting to know what "more stability" means. This, however, requires more specific assumptions about payoff functions as this has be done for instance in the theoretical context by Finus/Rundshagen (2001) or in the empirical context by Eyckmans/Finus (2003) for some of the coalition games we discussed here. Of course, one could argue that more is always better than less stability if one assumes that additional equilibria in one game compared to an other game are only chosen by players if they lead to higher aggregate welfare and lower global emissions. However, this would probably be a too simple view of the problem.

## References

Barrett, S. (1994a), The Biodiversity Supergame. "Environmental and Resource Economics", vol. 4, pp. 111-122.

Barrett, S. (1994b), Self-Enforcing International Environmental Agreements. "Oxford Economic Papers", vol. 46, pp. 804-878.

Barrett, S. (1997), Heterogeneous International Agreements. In: C. Carraro (ed.), International Environmental Negotiations: Strategic Policy Issues. Edward Elgar, Cheltenham, pp. 9-25.

Barrett, S. (1999), A Theory of Full International Cooperation. "Journal of Theoretical Politics", vol. 11, pp. 519-541.

Barrett, S. (2000), Consensus Treaties. Preliminary Draft, Johns Hopkins University, Washington D.C.

Bauer, A. (1992), International Cooperation over Greenhouse Gas Abatement. Mimeo, Seminar für empirische Wirtschaftsforschung, University of Munich, Munich.

Bloch, F. (1997), Non-Cooperative Models of Coalition Formation in Games with Spillovers. In: Carraro, C. and D. Siniscalco (eds.), New Directions in the Economic Theory of the Environment. Cambridge University Press, Cambridge, ch. 10, pp. 311-352.

Botteon, M. and C. Carraro (1997), Burden-Sharing and Coalition Stability in Environmental Negotiations with Asymmetric Countries. In: Carraro, C. (ed.), International Environmental Negotiations: Strategic Policy Issues. Edward Elgar, Cheltenham et al., ch. 3, pp. 26-55.

Carraro, C. (2000), Roads towards International Environmental Agreements. Siebert, H. (ed.), The Economics of International Environmental Problems, Mohr Siebeck, Tübingen, pp. 169-202.

Carraro, C., C. Marchiori and S. Oreffice (2001) Endogenous Minimum Participation in International Environmental Treaties. Mimeo, Fondazione Eni Enrico Mattei.

Carraro C. and C. Marchiori (2002), Stable Coalitions. Fondazione Eni Enrico Mattei, Working Paper No. 5.2002. Forthcoming in Carraro, C. (ed.), Endogenous Formation of Economic Coalitions, Edward Elgar, Cheltenham, UK.

Carraro, C. and D. Siniscalco (1993), Strategies for the International Protection of the Environment. "Journal of Public Economics", vol. 52, pp. 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", vol. 2, pp. 279-293.

Chander, P. and H. Tulkens (1997), The Core of an Economy with Multilateral Environmental Externalities. "International Journal of Game Theory", vol. 26, pp. 379-401.

Ecchia, G. and M. Mariotti (1998), Coalition Formation in International Environmental Agreements and the Role of Institutions. "European Economic Review", vol. 42, pp. 573-582.

Endres, A. (1997), Negotiating a Climate Convention - The Role of Prices and Quantities. "International Review of Law and Economics", vol. 17, pp. 201-224.

Endres, A. and M. Finus (1998), Renegotiation-Proof Equilibria in a Bargaining Game over Global Emission Reductions - Does the Instrumental Framework Matter? In: Hanley, N. and H. Folmer (eds.), Game Theory and the Global Environment. Edward Elgar, Cheltenham et al., ch. 7, pp. 135-164.

Eyckmans, J. and M. Finus (2003), Coalition Formation in a Global Warming Game: How the Design of Protocols Affects the Succees of Environmental Treaty-Making. Preliminary Draft, University of Hagen.

Eyckmans, J. and H. Tulkens (1999), Simulating with RICE Coalitionally Stable Burden Sharing Agreements for the Climate Change Problem. CES ifo Working Papers Series, No. 228, Munich.

Finus, M. (2001), Game Theory and International Environmental Cooperation. Edward Elgar, Cheltenham.

Finus, M. (2002), New Developments in Coalition Theory: An Application to the Case of Global Pollution. Forthcoming in Rauscher, M. (ed.), "The International Dimension of Environmental Policy", Kluwer, Dordrecht, Holland.

Finus, M. (2003), Stability and Design of International Environmental Agreements: The Case of Transboundary Pollution. Forthcoming in Folmer, H. and T. Tietenberg (eds), International Yearbook of Environmental and Resource Economics, 2003/4, Edward Elgar, Cheltenham, UK.

Finus, M. and B. Rundshagen (1998a), Toward a Positive Theory of Coalition Formation and Endogenous Instrumental Choice in Global Pollution Control. "Public Choice", vol. 96, pp. 145-186.

Finus, M. and B. Rundshagen (1998b), Renegotiation-Proof Equilibria in a Global Emission Game When Players Are Impatient. "Environmental and Resource Economics", vol. 12, pp. 275-306.

Finus, M. and B. Rundshagen (2001), Endogenous Coalition Formation in Global Pollution Control. A Partition Function Approach. Working Paper No. 307, University of Hagen. Revised version forthcoming in Carraro, C. (ed.), Endogenous Formation of Economic Coalitions, Edward Elgar, Cheltenham, UK.

Finus, M. and B. Rundshagen (2003), A Non-cooperative Foundation of Core-Stability in Positive Externality NTU-Coalition Games. Preliminary Draft, University of Hagen.

Finus, M. and S. Tjøtta (2002), The Oslo Protocol on Sulfur Reduction: The Great Leap Forward? "Journal of Public Economics", forthcoming.

Folmer, H., and P. van Mouche (2000), Transboundary Pollution and International Cooperation. In: Tietenberg T. and H. Folmer (eds), The International Yearbook of Environmental and Resource Economics, Cheltenham, UK and Brookfield, US: Edward Elgar, ch. 6, pp. 231-267.

Germain, M., H. Tulkens and A. de Zeeuw (1998), Stabilité stratégique en matière de pollution internationale avec effet de stock: Le cas linéaire. "Revue Économique", vol. 49, pp. 1435-1454.

Hart, S. and M. Kurz (1983), Endogenous Formation of Coalitions. "Econometrica", vol. 51, pp. 1047-1064.

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

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

Jeppesen, T. and P. Andersen (1998), Commitment and Fairness in Environmental Games. In: Hanley, N. and H. Folmer (eds.), Game Theory and the Environment. ch. 4, pp. 65-83, Edward Elgar, Cheltenham et al.

Mäler, K.-G. (1994), Acid Rain in Europe: A Dynamic Perspective on the Use of Economic Incentives. In: van Ierland, E. C. (ed.), International Environmental Economics. Developments in Environmental Economics 4. Elsevier, Amsterdam, pp. 351-72.

Rubio, S. and A. Ulph (2001), A Simple Dynamic Model of International Environmental Agreements with a Stock Pollutant. Preliminary Version, University of Southampton.

Rundshagen, B. (2002), On the Formalization of Open Membership in Coalition Formation Games. Working Paper No. 318, University of Hagen.

Stähler, F. (1996), Reflections on Multilateral Environmental Agreements. In: Xepapadeas, A. (ed.), Economic Policy for the Environment and Natural Resources: Techniques for the Management and Control of Pollution. Edward Elgar, Cheltenham and Brookfield, ch. 8, pp. 174-196.

Tulkens, H. (1998), Cooperation versus Free-Riding in International Environmental Affairs: Two Approaches. In: Hanley, N. and H. Folmer (eds.), Game Theory and the Environment. Edward Elgar, Cheltenham et al., ch. 2, pp. 30-44.

Yi, S.-S. (1997), Stable Coalition Structures with Externalities. "Games and Economic Behavior", vol. 20, pp. 201-237.

Yi, S.-S. and H. Shin (1995), Endogenous Formation of Coalitions in Oligopoly. Mimeo, Department of Economics, Dartmouth College.

Yi, S.-S. and H. Shin (2000), Endogenous Formation of Research Coalitions with Spillovers. "International Journal of Industrial Organization", vol. 18, pp. 229-256.