Fondazione Eni Enrico Mattei

Endogenous Coalition Formation in Global Pollution Control

Michael Finus* and Bianca Rundshagen*

NOTA DI LAVORO 43.2001

JUNE 2001

Coalition Theory Network

*University of Hagen, Germany

Fondazione Eni Enrico Mattei
Corso Magenta, 63, 20123 Milano, tel. +39/02/52036934 – fax +39/02/52036946
E-mail: letter@feem.it
C.F. 97080600154

1. Introduction

Concern about transboundary and global pollution problems ranks prominently on the agenda of international politics and has led to the signature of several international environmental agreements (IEAs), as for instance the Oslo Protocol on sulfur reduction in Europe in 1994, the Montreal Protocol on the depletion of the ozone layer in 1987 and the Kyoto-Protocol on the reduction of greenhouse gases in 1997. This concern is also reflected in numerous recent papers on the formation of coalitions in international pollution control since the appearance of Barrett (1991), Bauer (1992) Black/Levi/de Meza (1992), Carraro/Siniscalco (1991), Chander/Tulkens (1991), Hoel (1992) and Tulkens (1979). The *fundamental assumption* of all models is that IEAs must be *self-enforcingly* designed since there is no international agency that can establish binding agreements (Endres 1996). The main problem analyzed by these models is *free-riding* in international pollution control. There are two types of free-riding which negatively affect the success of an IEA (Finus 2000, ch. 2). The *first type of free-riding* is the incentive of a country remaining a non-signatory (or to choose a low abatement level), benefiting from the (higher) abatement efforts of (other) signatories. The *second type of free-riding* relates to the incentive of a signatory to violate the spirit of an agreement. Through free-riding a country can reduce its abatement effort substantially, though environmental quality will only be affected marginally. Thus a country can (temporarily) net a free-rider gain.

Though models differ with respect to the specification of the utility function of governments and with respect to the stability concept they employ, they can be classified in two groups.

Dynamic Game Models

The first group may be called *dynamic game models* (DG-models). These models assume an infinitely repeated game where governments agree on some contract in the first stage that has to be enforced in subsequent stages by using credible threats (Barrett 1994a, b, Finus/Rundshagen 1998 and Stähler 1996). Credibility is defined in these papers in terms of renegotiation-proof equilibrium. Whereas Barrett and Finus/Rundshagen consider only single deviations from the obligations of an IEA, Stähler also considers multilateral deviations, however, he restricts the number of countries to three which allows to draw only limited conclusions for transboundary and global environmental problems. Barrett analyzes the condition for stability of a grand coalition targeting at a globally optimal solution. He finds that stability may be jeopardized even in a supergame framework and even if agents are almost perfectly patient. Therefore, Finus/Rundshagen extend his work considering the formation of subcoalitions and of less ambitious abatement targets. They show that the allocation of the abatement burdens crucially affects the success of IEAs, a grand coalition is unlikely to form and that a subcoalition may achieve more than the grand coalition.

Reduced Stage Game Models

The second group of models may be termed *reduced stage game models* (RSG-models). Two variants may be distinguished: a) models applying the concept of *internal&external stability* and b) those applying the concept of the *core* to determine the equilibrium coalition structure. Both variants model coalition formation as a two-stage game where countries decide in the first stage on coalition formation and in the second stage countries choose their emission (abatement) levels and how they distribute the gains from cooperation.

Variant a): Internal&External Stability

Up to now, all models of variant a have assumed that in the first stage countries have only the choice between acceding to an IEA or to remain a non-signatory where non-signatories play as singletons. In the second stage, it has been assumed that signatories jointly maximize the coalition´s welfare and that non-signatories maximize their individual welfare. Signatories and non-signatories play non-cooperatively against each other. Either signatories and non-signatories choose simultaneously (Nash-Cournot assumption, e.g., Carraro/Siniscalco 1991 and 1993, Bauer 1992 and Hoel 1992) or sequentially (Stackelberg assumption, e.g., Barrett 1994b). For the distribution of benefits, three typical assumptions have been made. i) no distribution (Bauer 1992), ii) distribution according to the Nash-bargaining solution (Botteon/Carraro 1997 and 1998) and iii) distribution according to the Shapley value (Barrett 1997b and Botteon/Carraro 1997 and 1998). If countries are assumed to be ex-ante symmetric[1], all signatories receive the same payoff and a redistribution is obviously not necessary.[2] In the case of asymmetric countries, assumptions ii and iii are popular allocation rules, though other rules (as for instance the Kalai-Smorodinski-solution, Kalai/Smorodinski 1975, or the

---

1 Ex-ante symmetric refers to the assumption that all countries have the same payoff function though they may receive different payoffs depending whether they become a signatory or remain a non-signatory (or in the general context which coalition they join).

2 In the case of symmetric countries the Nash-bargaining, the Shapley value and probably any other bargaining rule of cooperative game theory implies no redistribution of the initial payoff allocation. However, strictly speaking, despite the assumption of symmetric countries any asymmetric allocation of payoffs cannot be ruled out a priori, though it would probably require some additional motivation that should be endogenous to the model.

proportional solution, Kalai 1977) are also conceivable.[3] Since payoffs are received at the end of the second stage and since the rules of the game (choice of emission levels, distribution of payoffs and procedure of coalition formation) are exogenously fixed, the two stages can be reduced (via backwards induction) to one stage that simplifies computations a lot. Coalition formation can be studied with the help of a partition function (see section 2), which contains all information about the payoffs a country receives in different coalition structures. In equilibrium, *internal stability* implies that no signatory has an incentive to leave the coalition; *external stability* implies that no non-signatory wants to accede to the agreement (see section 3.2.2 for a formal definition).

A key result of these models is that, generally, the number of signatories falls short of the grand coalition and the equilibrium coalition is rather small. Moreover, whenever cooperation would be needed most from a global point of view, the coalition achieves only little.[4] This result is also reminiscent to the dynamic game models of Barrett (1994a, b) and Finus/Rundshagen (1998).

Variant b): Core

In models of variant b it is analyzed in the first stage whether there is an incentive for one country or a group of countries to deviate from some coalition structure. In the second stage, it is assumed that coalitions choose an emission vector, which maximizes the aggregate payoff to the coalition and that in the case of asymmetric countries a transfer scheme is established. If a country or some countries deviate, it is assumed that the remaining players split up into singletons and i) minimax the deviating player(s) (α-core), ii) maximin the deviating player(s) (β-core) or iii) play a Nash strategy (γ-core). Simply speaking, a coalition structure is an equilibrium if no member can achieve a higher payoff in any other coalition structure. The key result of these models (Chander/Tulkens 1992, 1995, 1997) is that by choosing a cleverly designed transfer scheme (which resembles that of Kaneko´s ratio equilibrium), the grand coalition establishing the socially optimal emission vector can be sustained as an equilibrium.

---

3 Botteon/Carraro (1997 and 1998) nicely demonstrated the importance of the allocation rule for the number of signatories, for the composition of signatories and for the success of an IEA through a simulation exercise.

4 That is, in equilibrium, global welfare generated by the coalition is close to that in the Nash equilibrium and the gap between the coalition equilibrium and the social optimum is large. This result is demonstrated in Finus (2000b, ch. 13) for a large number of models belonging to variant a.

Evaluation and Comparison

The advantage of DG-models is that they capture both dimensions of free-riding. In contrast to RSG-models, these models also consider the possibility that countries may join an agreement but do not fully comply with the terms of the agreement. These models have been criticized for their assumption of an infinite time horizon and that they specify threat strategies in terms of a temporary expansion of emissions. It has been argued that political agents act in a finite time horizon and that emissions are not used in reality to punish free-riders.[5] Moreover, though these models would allow to account for simultaneous coalitions, given the nature of these models, the study of simultaneous coalitions would be a complex undertaking and would have to rely on simulations.

The advantage of RSG-models is related to their simplicity, though all share the disadvantage of capturing only one dimension of the free-rider problem, namely, that a country free-rides on the abatement efforts of other countries.[6] The second dimension of a country acceding to a coalition but violating the agreement is not captured. By assumption countries instantaneously reoptimize their strategy in the second stage and therefore a temporary free-rider gain is ruled out.[7] A particular drawback of those RSG-models, which apply the stability concept of internal&external stability, is that they assume exogenously that there will be only one non-trivial[8] coalition (signatories). Of course, this assumption simplifies computations tremendously, however, a priori it is not clear whether the co-existence of several coalitions could not also be an equilibrium. This concern is taken up by those RSG-models that apply the stability concept of the core. This variant of models allows for the co-existence of several coalitions. An other advantage is that results do not have to rely on simulations (as this is the case for almost all previously mentioned models). Moreover, the core concept allows studying stability of an

---

5 See Finus (2000a, b) for a discussion of these issues and possible arguments of defense.

6 In the models applying the concept of internal&external stability free-riding implies to remain a non-signatory. In the reduced stage models applying other concepts (and which allow for the possibility of the co-existence of several coalitions), free-riding implies that a country belongs to a coalition that contributes less to joint abatement than other coalitions. See sections 2 and 3.

7 This allows for two interpretations of an equilibrium in RSG-models: it has to be assumed that either a) countries comply with the terms of an IEA once they have acceded or that b) the violation of an IEA is immediately discovered. The first interpretation is not in accordance with the empirical evidence on the compliance of numerous IEAs (see the literature cited in Finus (2000a, ch. 2); the second interpretation requires an optimistic view as to the possibilities of monitoring and the flexibility of altering strategies.

8 A non-trivial coalition means a coalition of at least two countries.

IEA where the members gradually adjust to the targets of the treaty (e.g., Chander/Tulkens 1991 and 1992, Germain/Toint/Tulkens 1996 and Germain/Toint/Tulkens/de Zeeuw 1998). It thus allows capturing an important feature of actual IEAs where signatories achieve their long-term targets not in one step but in a step-by-step fashion. The disadvantage of this variant of models is that from a positive point of view they do not contribute very much to explain the formation of IEAs in reality: none of the existing IEAs constitutes a grand coalition and transfers played a neglectable role if at all in the past. Due to the assumption that a) there are no problems to administer and to enforce transfers and b) if a deviation occurred the remaining players would resolve the coalition, this optimistic result is obtained, rendering the co-existence of coalitions only a theoretical possibility which never materializes.

Our model framework belongs to the group of RSG-models and relies on work of Bloch (1997) and Yi (1997). Like the core concept, our framework allows for the possibility of several coalitions. However, equilibrium coalition structures are determined by applying "new concepts" to study coalition formation. Some of these concepts, which we call coalition formation games, have been applied to problems of industrial economics and international trade whereas some others have not been applied yet (see the literature cited in Finus 2001, ch. 15). Our main concern is to compare the equilibrium coalition structures in global pollution control under different coalition formation games and to evaluate these games with respect to their theoretical properties and how they capture features of the formation of existing IEAs. Since these concepts imply different rules according to which coalitions form, they also have a normative dimension. For instance, one key issue is whether membership in an IEA should be based on unanimous agreement by all signatories (exclusive membership) or whether all countries should be allowed to accede to an IEA if they wish do so (open membership).

It should be pointed out right at beginning that for the complexity of allowing for the possibility of the co-existence of several coalitions we have to pay a price in terms of making the simplifying assumption of ex-ante symmetric countries. Thus, the issue of redistributing the benefits of a coalition and the choice of the abatement target is trivially solved.[9] Moreover, for some concepts it is necessary to assume a specific payoff function (though simulations are not necessary) to derive interesting conclusions. However, it should be kept in mind, that all models mentioned above share this disadvantage, except the reduced stage game models of variant b.

---

[9] There is no redistribution and countries choose that emission level which maximizes a coalition´s aggregate payoff (which also maximizes a country´s payoff).

In what follows, we lay out the model framework, present definitions and derive some general results in section 2, which form the bases of determining the equilibrium coalition structures under various coalition formation games in section 3. In section 4 we compare the equilibrium coalition structures, draw conclusions and evaluate the concepts theoretical and practical potential to explain the formation of IEAs and point to issues of future research.

2. Model

2.1 Definitions

We consider a RSG comprising *two stages* where in the *first stage* countries form coalitions and in the second stage coalitions choose their optimal strategy vector. The details of the first stage are summarized in the *rules of a coalition formation game*. We postpone the details until section 3.

In the second stage, we assume that the *members of a coalition maximize aggregate payoffs to the coalition and play a Nash-Cournot strategy against other coalitions*. Since we assume *ex-ante symmetric countries*, this implies that those countries that belong to the same coalition choose the same emission level and receive the same payoff. We therefore do not have to consider transfers among coalition members in the second stage. Transfers between coalitions are also ruled out.

Assumption 1: Rules of the Second Stage

Countries are assumed to be ex-ante symmetric, and to maximize a coalition's aggregate payoff. There are no transfers.

If the rules of the second stage are ex-ante specified, the two stages can be reduced to one stage. All relevant information on which the decision in the first stage is based can then be compactly summarized in the per-membership partition function (Bloch 1997).

Definition 1: Equilibrium Valuation or Per-Membership Partition Function

Let $c = \{c_1, ..., c_M\}$ denote a coalition structure with M coalitions, $c \in C$ where C denotes the set of coalition structures, $c_j \cap c_k = \varnothing \; \forall j \neq k$, $c_1 \cup ... \cup ... \cup c_M = I$, then an equilibrium valuation is a mapping which associates to each coalition structure $c \in C$ a vector of individual payoffs $\pi(c) = \{\pi_1(c_j, c), ..., \pi_N(c_k, c)\}$ where $\pi(c) \in \Pi(C)$ is the set of payoffs which results from the maximization of players according to a particular rule, a given sharing rule of the gains from cooperation and a given coalition structure. The first argument in $\pi_i(c_i, c)$ refers to the coalition to which country i belongs, the second to the particular coalition structure.

For Assumption 1 and strictly concave payoff functions (see section 2.2), there is a unique optimal strategy vector for each possible coalition structure in the second stage. Thus, the set of equilibrium valuations is uniquely defined.

The coalition formation may be summarized as follows:

Definition 2: Coalition Formation Game

In the coalition formation game a country decides, based on the per-membership partition function, on the membership according to the rules of a coalition formation game. The partition function is determined by the rules of the second stage in the game.

Note that due to the assumption of symmetry, notation simplifies. Thus, instead of writing for example $C^*=\{c^{1*}, c^{2*}, c^{3*}\}$ where $c^{1*}=\{\{1, 2\}, \{3\}\}$, $c^{2*}=\{\{1\}, \{2, 3\}\}$, $c^{3*}=\{\{1, 3\}, \{2\}\}$ (implying all coalition structures that two countries can form and where one country remains a singleton are equilibrium coalition structures), we write $C^*=\{(2, 1)\}$ or $c^*=(2, 1)$ where the entries indicate the coalition sizes and the asterisks equilibrium coalition structures.[10] For the subsequent notation it will prove helpful to order coalitions according to their size. That is, $c=(c_1, ..., c_M)$ where $c_1 \geq ... \geq c_M$.

An important definition to compare different coalition structures is that of coarsening and concentration (e.g., Yi (1997), p. 205 and Bloch (1997), p. 334).

Definition 3: Coarsening of a Coalition Structure

$c=(c_1, c_2, ..., c_M)$ is a coarsening of $c'=(c'_1, c'_2, ..., c'_{M'})$, $M<M'$ if and only if there is a sequence of coalition structures $c^1=(c^1_1, c^1_2, ..., c^1_{M(1)})$, $c^2=(c^2_1, c^2_2, ..., c^2_{M(2)})$, ..., $c^R=(c^R_1, c^R_2, ..., c^R_{M(R)})$ with $M(i-1)=M(i)-1$ for all i = 2, ..., R such that 1) $c=c^1$ and $c'=c^R$ and 2) $c^{r-1}=c^r \backslash \{c^r_i, c^r_j\} \cup \{c^r_i+c^r_j\}$ for some $i, j \in \{1, ..., M(r)\}$ and for all r = 2, ..., R.

A coalition structure c is coarser than a coalition structure c' if c can be obtained by merging coalitions in c'. For example coalition structure (6, 5) is coarser than coalition structure (5, 5, 1) since (6, 5) can be obtained by merging coalitions 5 and 1 in coalition structure (5, 5, 1). However, many coalition structures cannot be compared under coarsening as for instance (5, 5) and (6, 4). Then a comparison may be possible under the criterion of concentration.

[10] Thus c_1 denotes a coalition within a coalition structure and c^1 a particular coalition structure.

Definition 4: Concentration of a Coalition Structure

$c = (c_1, c_2, ..., c_M)$ is a concentration of $c' = (c'_1, c'_2, ..., c'_{M'})$, $M \leq M'$ if and only if there is a sequence of coalition structures $c^1 = (c^1_1, c^1_2, ..., c^1_{M(1)})$, $c^2 = (c^2_1, c^2_2, ..., c^2_{M(2)})$, ..., $c^R = (c^R_1, c^R_2, ..., c^R_{M(R)})$ such that 1) $c = c^1$ and $c' = c^R$ and 2) $c^{r-1} = c^r \setminus \{c^r_{i(r)}, c^r_{j(r)}\} \cup \{c^r_{i(r)} + 1, c^r_{j(r)} - 1\}$, $c^r_{i(r)} \geq c^r_{j(r)}$ for some i(r), j(r) = 1, ..., M(r) and for all r=2, ..., R.

That is, c is a *concentration* of c' if one can obtain c by moving one member at a time from a coalition in c' to another coalition of equal or larger size. Through this process, coalitions in c' *may* sequentially be dissolved. For instance (6, 5) is a concentration of (5, 5, 1) since the singleton coalition is dissolved and this player joins a larger coalition of size 5. However, (6, 4) is also a concentration of (5, 5), though no coalitions are dissolved. Unfortunately, however, also concentration does not allow for a complete ordering of coalition structures. For instance, (4, 3) and (5, 1, 1) cannot be ranked under concentration.

From the definitions it follows that every coalition c which is coarser than a coalition c' implies that c is a concentration of c', however, the opposite is not true. A formal proof of this relation is provided in Yi (1997).

2.2 Properties of the Global Pollution Game

In its most general form, the payoff function of the global emission game may be written as follows:

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

where we assume $\beta' > 0$ for all $e_i < e_i^{max}$, $\beta'' < 0$ for all $e_i > 0$, $\phi' > 0$ and $\phi'' \geq 0$ for all $\sum e_j > 0$. That is, benefits from emission (in the form of consumption and production of goods) increase in emissions at a decreasing rate. Damages increase in global emissions at a constant or increasing rate.

Unfortunately, as pointed out in the Introduction, the amount of conclusions, which can be derived for the general function [1], is very limited. Therefore, we consider two additional examples that have widely been used in the literature on coalition formation.

$$[2] \qquad \pi_i = b(de_i - \frac{1}{2}e_i^2) - c(\sum_{j=1}^{N} e_j)$$

$$[3] \qquad \pi_i = b(de_i - \frac{1}{2}e_i^2) - \frac{c}{2}(\sum_{j=1}^{N} e_j)^2$$

Whereas payoff function [2] assumes constant marginal damages, payoff function [3] assumes linear marginal damages. This implies orthogonal reaction functions for payoff function [2]. That is, the slope of the reaction function is zero. In contrast, for payoff function [3] reaction functions are downward sloping in emission space with a slope greater than –1 and less than zero.[11] Thus payoff function [3] exhibits a more interesting pattern than [2] with respect to the interaction of agents. However, also for payoff function [3] only a limited amount of general properties can be established which allow making precise predictions about coalition formation. Therefore, in most parts of the paper we will illustrate results based on payoff function [2].

Following Yi (1997), the global emission game maybe viewed as a *positive externality game* if a strategy is seen as emission reduction from some status quo.[12] That is, if a country or a coalition reduces emissions, all other countries or coalitions benefit from abatement efforts as well. In the standard framework without considering coalition formation this is an immediate implication of $\partial \pi_i / \partial e_j < 0$. In the context of coalition formation this fact also appears but also other facets of it. To demonstrate those, it is helpful to look first at emissions of countries belonging to different coalitions and at global emissions resulting from different coalition structures.

Proposition 1: Emissions

a) Let emissions of a member of a coalition c_i be denoted by e_i and of a coalition c_j by e_j, then for any coalition structure c we have $e_i > e_j$ iff $c_i < c_j$.

b) Let coalition structure c' be coarser than coalition structure c and denote total emissions by e^T, then $e^T(c') < e^T(c)$.

c) Let coalition structure c' be more concentrated than coalition structure c, then $e^T(c') < e^T(c)$ for payoff functions [2] and [3].

---

11 Note that the following results do not depend on the exact specification of functions [2] and [3]. For the results it is only important that payoff function [2] implies constant marginal damages and payoff function [3] linear marginal damages. In the context of coalition formation payoff function of type [2] has been used for instance by Barrett (1994b, 1997a), Bauer (1992), Botteon/Carraro (1997) and Hoel (1992), payoff function of type [3] by Barrett (1994b), Carraro/Siniscalco (1991), Finus/Rundshagen (1998) and Stähler (1996). Only for the concept of the core Chander/Tulkens (1995, 1997) were able to produce results based on a general payoff function of type [1].

12 In our context it seems obvious to define the status quo as the Nash equilibrium with only singleton coalitions.

Proof: a) follows from the first order condition of a member of a coalition c_k which is given by $\beta'(e_k) = c_k\phi'(e^T)$. Since for any e^T $c_k\phi'(e^T)$ increases in c_k and $\beta'' < 0$, e_k increases in the size of the coalition c_k. b) is demonstrated by contradiction. Assume $e^{T(1)} < e^{T(2)}$ would be true where $e^{T(1)}$ are global emissions before and $e^{T(2)}$ after coalitions c_i and c_j have merged. Moreover, assume that there is a third coalition c_k that is not involved in the merger. Then for members of coalitions

k: $\beta'(e_k^{(1)}) = c_k\phi'(e^{T(1)}) < c_k\phi'(e^{T(2)}) = \beta'(e_k^{(2)}) \Rightarrow e_k^{(1)} > e_k^{(2)}$
i (and j): $\beta'(e_i^{(1)}) = c_i\phi'(e^{T(1)}) < (c_i + c_j)\phi'(e^{T(1)}) < (c_i + c_j)\phi'(e^{T(2)}) = \beta'(e_i^{(2)}) \Rightarrow e_i^{(1)} > e_i^{(2)}$
hold which obviously violates the initial assumption of $e^{T(1)} < e^{T(2)}$. c) follows from routine computations which shows that global emissions are given by

payoff function [2]: $e^T = Nd - \frac{c}{b}\sum_{i=1}^{M} c_i^2$ and payoff function [3]: $e^T = \frac{Nd}{1 + \frac{c}{b}\sum_{i=1}^{M} c_i^2}$

respectively and where $\sum_{i=1}^{M} c_i^2$ increases through concentration since, assuming $c_i \le c_j$, $[(c_i - 1)^2 + (c_j + 1)^2 + \sum_{k \ne i,j} c_k^2] - \sum_k c_k^2 = 2(1 + c_j - c_i) > 0$ holds (**Q.E.D.**).

Proposition 1a is an immediate implication of Assumption 1: coalition members maximize the aggregate payoff to their coalition. The larger a coalition, the more do their members care about the negative impact their emissions exhibit on other countries. Consequently, global emissions decrease if coalitions merge and form larger coalitions as stated in Proposition 1b. Only for coalition structures, which cannot be ranked under coarsening, such a general conclusion is not possible. However, as Proposition 1c demonstrates, global emission decrease for payoff functions [2] and [3] if coalition structures become more concentrated. An immediate implication of Proposition 1 is the following corollary.

Corollary 1: Global Emissions

The grand coalition produces the lowest global emissions, the degenerated coalition structure consisting of singletons i produces the highest global pollution.

Proof: Follows from Proposition 1b and the facts that a) N is coarser than c if c≠(N) and b) c is coarser than (1, 1,..., 1) (if c≠(1, 1, ..., 1)) for each coalition structure c (**Q.E.D.**).

We can now turn to the facets of a positive externality game in the context of coalition formation. One facet is that members of small coalitions enjoy a higher payoff than members of large coalitions for any given coalition structure.

C_1: $\pi_j(c_j, c) < \pi_i(c_i, c)$ *iff* $c_i < c_j$

Since all countries suffer equally from emissions but members of smaller coalitions choose higher emissions in equilibrium (and therefore have higher benefits), members of smaller coalitions are better off than members of larger coalitions. A somewhat more specific though still very general feature of the positive externality game is the following.

C_2:

a) $\pi_i(c_i, c) < \pi_i(c_i, c')$ *where* $c_i \subset c, c'$ *and* c' *is coarser than c.*

b) $\pi_i(c_i, c) < \pi_i(c_i, c')$ *where* $c_i \subset c, c'$ *and* c' *is more concentrated than c.*

Whereas condition a) claims that coalitions which are not involved in a merger are better off after the merger, condition b) claims that coalitions which are not involved in a concentration are better off if other coalitions form a more concentrated coalition structure. Obviously, condition b) is stronger than a). C_2 a) and b) are implications of Proposition 1 b) and c), respectively: global emissions decrease through coarsening (concentration) which has a positive effect on outsiders.[13]

Conditions 1 and 2 reveal a typical feature of a positive externality game, namely that of free-riding. Smaller coalitions and those coalition members, which are not involved in forming larger coalitions aiming at reducing global damages, benefit from the abatement efforts of the more active players. Other features, though more specific, are also related to this problem. The next two conditions deal with the effect on members of an "old" and "new" coalition if a member leaves his old coalition i to join a new coalition j.

C_3: $\pi_i(c_i, c) < \pi_i(c_i \setminus \{k\}, c')$ *where* $c' = c \setminus \{c_i, c_j\} \cup \{c_j \cup \{k\}, c_i \setminus \{k\}\}$, $2 \le c_i \le c_j$.

If a member of the coalition i leaves his coalition to join a larger or equal-sized coalition j, the members of the old coalition are better off. The old members benefit from the increased abatement efforts of the new and larger coalition (payoff function [2] and [3]) while they keep their efforts constant (payoff function [2]) or reduce their effort (payoff function [3]).

For payoff functions [2] and [3] it turns out that the free-rider incentive can be compactly summarized as follows.

---

[13] Outsiders increase their emissions after the concentration (see the proof of Proposition 1) and hence receive higher benefits and lower damages.

C_4: $\pi_k(c_i, c) > \pi_k(c_j \cup \{k\}, c')$ where $c' = c \setminus \{c_i, c_j\} \cup \{c_j \cup \{k\}, c_i \setminus \{k\}\}$ if $c_j \geq c_i \geq 2$.

If a member k of the coalition i leaves his coalition to join a larger or equal-sized coalition j, then the deviator becomes worse off. From C_4 it also follows that joining a coalition that is smaller by at least two members is profitable. Condition C_4 stresses that - once there is some (minimal) amount of concentration - single members have only a marginal effect on reducing global damages by joining larger coalitions but their benefits decrease substantially through increased abatement efforts. With respect to a merger the following condition can be established.

C_5: $\pi_j(c_j, c_i, c) < \pi_j(c_j \cup c_i, c)$ and $\pi_i(c_i, c) < (\geq) \pi_i(c_j \cup c_i, c')$ if $c_j - 2\ \ c_i < (\geq)\ 0$ where $c' = c \setminus \{c_i, c_j\} \cup \{c_j \cup c_i\}$ and $c_j \geq c_i$.

The smaller coalition i must be at least half the size of the bigger coalition to be not worse off after the merger of coalition i and j. For $c_j = c_i = 1$ coalition c_i gains from a merger. The larger or equal-size coalition j always gains.

In order to compactly summarize our results in sections 3 and 4, we make the following assumption related to C_4 and C_5, which we assume to hold in the remainder of this paper.

Assumption 2: Indifference of Payoffs

If players are indifferent between being a member of a smaller or larger coalition, they join the larger coalition.

In the following, Assumption 2 implies for instance that if a singleton is indifferent between remaining a singleton and joining a coalition of size two and the larger coalition likes him to join, we assume he will do so (see conditions C_4 and C_5).[14]

Our discussion of conditions C_1 to C_5 may be summarized as follows.

Proposition 2: Incentive Structure of the Global Emission Game

The general payoff function [1] satisfies conditions 1 and 2a, payoff function [2] satisfies condition C_1 to C_5 and payoff function [3] conditions C_1 to C_4.

Proof: The statement with respect to the general payoff function has already been proved above. The statement with respect to payoff functions [2] and [3] is proved in Appendix 1 **(Q.E.D.)**.

---

14 We take this assumption from Ray/Vohra (1999). It reduces the amount of knife-edge equilibria but does not affect the fundamental results.

In section 3 it turns out that conditions C_1 to C_5 are essential for characterizing equilibrium coalition structures for most coalition games. For only a few games C_1, C_2 (strong version) and C_3 are sufficient to draw some conclusions, for some other games also C_4 is needed and for some games without C_5 almost nothing can be said. Since we find it interesting to compare equilibrium coalition structures between different coalition games, it is evident that we have to work with payoff function [2] to make progress. This is also evident when considering the subsequent properties that we will use in section 3.

The subsequent Proposition 3 will be helpful in that it allows drawing immediate conclusions with respect to global welfare if equilibrium coalition structures under different coalition games can be ranked according to concentration.

Proposition 3: Concentration and Global Welfare

Assume payoff function [2]. If c is a concentration of c′, global welfare is higher under c.

Proof: See Appendix 2 (**Q.E.D.**).

At a more general level, we find:

Proposition 4. Welfare in the Grand Coalition

The grand coalition produces the highest global welfare.

Proof: Follows obviously from $\max \sum_{i \in (N)} \pi_i \geq \max \sum_{i \in c_1} \pi_i + \max \sum_{j \in c_2} \pi_j + ... + \max \sum_{k \in c_M} \pi_k$ (**Q.E.D.**).

For the cartel formation, open-membership and the exclusive membership Δ-game the following definition and proposition will turn out to be useful.

Definition 5: Stand-alone Stability of a Coalition Structure

$c=\{c_1, ..., c_M\}$ is stand-alone stable iff $\pi_i(c_i, c) \geq \pi_i(\{i\}, c')$ where $c' = c \setminus c_i \cup \{c_i \setminus \{i\}, \{i\}\}$ $\forall i \in I$.

A coalition structure c is *stand-alone stable* if and only if no player finds it profitable to leave her coalition to be a singleton, holding the rest of the coalition structure constant. From this it follows immediately that *the degenerated coalition structure consisting only of singletons is stand-alone stable by definition*.

Proposition 5: Stand-alone Stability for Payoff Functions [2] and [3]

For payoff function [2] all coalition structures where no coalition comprises more than three coalition members is stand-alone stable. For payoff function [3] a coalition structure may only be stand-alone stable if no coalition comprises more than two coalition members.

Proof: See Appendix 3 (**Q.E.D.**).

Interestingly, assuming payoff function [3], the more concentrated a coalition structure is, the more likely it is that the stand-alone stability fails. That is, the more coalitions of size two have formed, the higher is the incentive to take a free-ride and to become a singleton (see Appendix 3).

To derive the equilibrium coalition structure under the exclusive membership Δ-game and the sequential move unanimity game but also to evaluate equilibrium coalition structures in generally, the following definition and proposition will turn out to be useful.[15]

Definition 6: Pareto-optimal Coalition Structures (POs)

A coalition structure c is Pareto-optimal if there is no other coalition structure c' *where at least one player is better off and no player is worse off, i.e.,* $\forall\ c' \neq c$ *with* $\pi_i(c_i', c') > \pi_i(c_i, c)$ *for some* $i\ \exists\ j \in I : \pi_j(c_j', c') < \pi_j(c_j, c)$. *This implies that there is no other coalition structure* c' *which weakly Pareto-dominates c.*

Proposition 6: Pareto-optimal Coalition Structures and Grand Coalition

The grand coalition is a Pareto-optimal coalition structure.

Proof: Let $c = (c_1, ..., c_M)$ where $c_1 \geq ... \geq c_M$. Then $\pi_1(c_1, c) \leq \pi_j(c_j, c)\ \ \forall\ \ c_j \leq c_1$ by condition C2a. Since $N \cdot \pi_i(N) \geq c_1 \cdot \pi_1(c_1, c) + \sum_{j=2}^{M} c_j \pi_j(c_j, c)$, $\pi_1(N) \geq \pi_1(c_1, c)$ holds. Hence the grand coalition is a PO (**Q.E.D.**).

Whereas for general payoff function [1] and also for payoff function [3] a more specific characterization beyond Proposition 6 is not possible, for payoff function [2] the entire set of POs can exactly be determined.

---

[15] It may be worthwhile to point out that our definition of a Pareto-optimal coalition structure assumes a fixed behavior of the coalition´s members. That is, a coalition maximizes the coalitions welfare and plays a Nash strategy against outsiders.

Proposition 7: Pareto-optimal Coalition Structures for Payoff Function [2]

For payoff function [2] the set of Pareto-optimal coalition structures, $C^{PO}(N)$*, is given by* $C^{PO}(N)=\{N\} \cup \{c=\{c_1,...,c_M\} \mid M \geq 2,\ c_1 \geq ... \geq c_M, \{c_1,..., c_M\} \backslash c_j \in C^{PO}(N-c_j)\ \forall j \in \{1, ..., M\},\ \pi_i(c_M, c) > \pi_i(N)\}$.

Proof: Follows from the following facts: a) $\{N\} \in C^{PO}(N)$ by Proposition 6. b) $c \in C^{PO}(N)$ => $\tilde{c} \in C^{PO}(\sum \tilde{c}_i)\ \forall \tilde{c} \subset c$. c) If $c \in \{N\} \cup \{c=\{c_1,...,c_M\} \mid M \geq 2,\ c_1 \geq ... \geq c_M,\ \{c_1,..., c_M\} \backslash c_j \in C^{PO}(N-c_j)\ \forall j \in \{1, ..., M\},\ \pi_i(c_M, c) > \pi_i(N)\}$, then the smallest coalition has no incentive to form the grand coalition[16] and no other coalition has an incentive to participate in forming any other coalition structure since each sub-coalition structure is a PO itself (**Q.E.D.**).

Table 1: Pareto-optimal Coalition Structures for Payoff Function [2]*

N	Pareto-Optima	N	Pareto-Optima
1	(1)	7	(7), (6, 1), (5, 2)
2	(2)	8	(8), (7, 1), (6, 2)
3	(3)	9	(9), (8, 1), (7, 2)
4	(4), (3, 1)	10	(10), (9, 1), (8,2), (7, 3)
5	(5), (4, 1)	11	(11), (10, 1), (9, 2), (8, 3)
6	(6), (5, 1)	12	(12), (11, 1), (10, 2), (9, 3), (8, 3, 1)

* Assumption 2 is assumed to hold.

Proposition 6 implies that $C^{PO}(N)$ are computed recursively. For N=1 to N=12, the set of Pareto-Optima are listed in Table 1. For instance for N=6, $C^{PO}(N)=\{(6),(5,1)\}$. (6) is the grand coalition. (5, 1) is a PO since (5, 1)\{5}=(1) is a PO (see N=1), (5, 1)\{1}=(5) (see N=5) is a PO and $\pi_i(1,(5,1)) > \pi_i(6,(6))$. For practical purposes of determining $C^{PO}(N)$, it is helpful to note that a necessary condition for a coalition structure $c=(c_1, c_2, ..., c_M)$, $c_1 \geq c_2 \geq ... \geq c_{M-1} \geq c_M$, to qualify as a PO is $c_i > 2c_{i+1}$ for any i=1...M-1 due to C_5. (Suppose the opposite is true, then there is an incentive for at least two coalitions to merge.)

16 Recall that due to C_1 the members of the smallest coalition derive the highest payoff in a given coalition structure. Due to Assumption 2, in a PO it is assumed that the smallest coalition will participate in the grand coalition if this leaves its members indifferent.

3. Coalition Structures

3.1 Introduction

We consider six different versions (coalition games) how the first stage of the coalition formation process can be modeled: 1) cartel formation game, 2) open membership game, 3) exclusive membership Δ-game, 4) exclusive membership Γ-game, 5) sequential move unanimity game and 6) equilibrium binding agreement game. These games can be structured with respect to two distinguishing features. The first feature is the *time dimension* of the coalition formation process. Either the formation process is modeled as a one-shot game or as a dynamic process. We choose this feature to group the games in this section. Coalition games 1 to 4 assume *simultaneous choice of membership* (section 3.2), whereas games 5 and 6 assume a *sequential choice of membership* (section 3.3). The second feature concerns the membership. In open membership type of games all players can freely accede to a coalition if they want. In exclusive membership type of games external players need the consent of the members of a coalition before they can join. Games 1 and 2 assume open membership, games 3 and 4 exclusive membership, and games 5 and 6 imply de facto exclusive membership, though this is not explicitly spelled out in the definition of these games.

In what follows we introduce the coalition games in subsections 3.2 and 3.3 and derive the equilibrium coalition structures. In section 4 we discuss the equilibrium coalition structures, evaluate the coalition games with respect to their theoretical consistency, their practical application and compare the equilibrium coalition structures among the different games.

3.2 Simultaneous Choice of Membership

3.2.1 Preliminaries

In order to select the equilibrium coalition structure(s) from equilibrium valuations, we need an equilibrium concept. For coalition games 1 to 4, we use the concepts Nash equilibrium (Nash 1950), strong Nash equilibrium (Aumann 1959) and coalition-proof Nash equilibrium (Bernheim/Whinston/Peleg 1987) that in our context may be defined as follows (see, e.g., Bloch 1997).

Definition 7: Nash and Strong Nash Equilibrium Coalition Structure (NE and SNE)

Let $G=\{I,\Sigma=\{\Sigma_i\}_{i\in I},\pi(c(\sigma))=\{\pi_i(c_i,c)\}_{i\in I}\}$ be the first stage of the coalition formation game with players $i \in I$, strategy vectors $\sigma \in \Sigma$ (proposals for coalitions), resulting coalition structures c and vectors of payoff functions π. Further, let $\tilde{C}(c^S,\sigma)$ be the set of coalition structures that a subgroup of countries c^S can induce if the remaining countries $j\in I\backslash c^S$

play σ_j. *For a fixed strategy vector* σ *define the reduced game for subgroup* c^S *as* $G^s_\sigma = \{c^S, \{\Sigma_i\}_{i\in c^S}, \{\pi_i(\tilde{c}_i, \tilde{C}(c^S,\sigma)\}_{i\in c^S}\}$. *Then* σ^* *is called a Nash equilibrium (strong Nash equilibrium) with the resulting Nash equilibrium (strong Nash equilibrium) coalition structure (NE, SNE)* c^* *if no singleton* $c^S = \{i\}$ *(no subgroup* c^S*) can increase his (at least one members') payoff (without reducing the payoff of any other member) by inducing another coalition structure. That is,*

$c^*(\sigma^*)$ *is a NE if* $\forall i \in I$ *and* $\forall \tilde{c} \in \tilde{C}(\{i\},\sigma^*)$: $\pi_i(c_i, c^*) \geq \pi_i(\tilde{c}_i, \tilde{c})$,

$c^*(\sigma^*)$ *is a SNE if no subcoalition* $c^S \subset I$ *can induce a coalition structure* $\tilde{c} \in \tilde{C}(c^S,\sigma^*)$ *with* $\pi_i(c_i, c^*) \leq \pi_i(\tilde{c}_i, \tilde{c})$ $\forall i \in c^S$ *and* $\pi_i(c_i, c^*) < \pi_i(\tilde{c}_i, \tilde{c})$ *for at least one* $i \in c^S$.[17]

Definition 8: Coalition-Proof Nash Equilibrium Coalition Structure (CPNE)

For $I = \{1\}$ σ_i *is a coalition-proof Nash equilibrium if and only if it is a Nash equilibrium. Assume that* $|I| = n > 1$ *and that coalition-proof Nash equilibrium coalition structures have been defined for all* $m < n$. *Then*

- σ *is self-enforcing if and only if for all* $c^S \subset I$, $c^S \neq I$, σ_{c^S} *is a coalition-proof Nash equilibrium of* G^s_σ.
- σ^* *is a coalition-proof Nash equilibrium of G with the coalition-proof coalition structure (CPNE)* c^* *if and only if it is self-enforcing and there does not exist another self-enforcing strategy* σ' *such that* $\pi_i(c_i(\sigma'), c(\sigma')) \geq \pi_i(c^*_i, c^*)$ $\forall$ $i \in I$ *and* $\pi_i(c_i(\sigma'), c(\sigma')) > \pi_i(c^*_i, c^*)$ *for at least one i.*[18]

Whereas a NE requires that a coalition structure is immune to deviations by single countries, a SNE also requires that deviations of any subgroup of countries are not beneficial to the deviators. From this it follows that any SNE is a NE too. That is, $C^{SNE} \subset C^{NE}$. Moreover, a *necessary* condition for a SNE is that it is a Pareto optimal coalition structure, i.e., $C^{SNE} \subset C^{PO}$. That is, a NE should not be weakly Pareto-dominated by an other NE. However, one has to be aware that, generally, not every weakly Pareto-undominated NE is a SNE, i.e., $C^{PO} \not\subset C^{SNE}$.

---

17 Originally, a SNE is defined as an equilibrium where no subcoalition can deviate such that the welfare of each member increases. We use a stronger version here to be consistent with Assumption 2. For example for payoff function [2] and N=3, according to our definition $c^* = (3)$ is a SNE, but not $c = (2,1)$, which could also be a SNE according to the original definition.

18 The previous footnote (with appropriate changes) applies to CPNE as well.

In contrast to a SNE, a CPNE must only be immune to deviations which are self-enforcing. Consequently, $C^{SNE} \subset C^{CPNE}$. For instance, $\pi_i(1,(3,1)) > \pi_i(4,(4)) > \pi_i(2,(2,2))$ is true for payoff function [2]. Hence, c=(2, 2) cannot be a SNE since this coalition structure is Pareto-dominated by grand-coalition. However, a country has an incentive to leave the grand coalition if the other countries remain in the coalition which may be the case depending on the assumptions of the coalition formation rules. Hence, c=(2, 2) might be a CPNE and in fact is one in the open-membership game and the exclusive membership Δ-game.
Since coalition-proofness considers self-enforcing deviations of subgroups of countries, the special case of single deviations is entailed in the definition and $C^{CPNE} \subset C^{NE}$. Thus, we have $C^{SNE} \subset C^{CPNE} \subset C^{NE}$.

Of course, one may wonder whether requiring a coalition structure to be Pareto-efficient with respect to the entire set of coalition structures is not an unduly restrictive condition. In particular since it turns out below that in some coalition games, as in many other games of economic interest, no SNE exists. However, the main weakness of the SNE-concept is that it does not impose any consistency requirement on deviations. A deviation is deemed feasible even though this deviation may be subject to further deviations. In contrast, coalition-proofness rules out such non-credible deviations. A coalition structure is only allowed to be challenged by self-enforcing deviations. The weakness of the CPNE-concept is that self-enforcing deviations are defined in a narrow sense: it only allows subsequent deviations by those players who deviated initially. Thus, only internal consistency but not external consistency of deviations is ensured by this concept. Though for future research one may want to develop a concept taking up this concern, we are not aware of any better concept presently. However, there is no doubt that such an extension would introduce a great complexity that may be difficult to handle.

3.2.2 Cartel Formation Game

The cartel formation game[19] has been widely applied in the environmental economics literature (e.g., Barrett 1994b, 1997a, b, Bauer 1992, Carraro/Siniscalco 1991, 1993 and Hoel 1992) because of its simplicity. Its roots go back to d´Aspermont et al. (1983) who used this setting to study cartel formation in an oligopoly. We study this game for reference reason, though it restricts the number of non-trivial coalitions to one. That is, it is exogenously assumed that there is one group of countries (signatories) which form a coalition and that all other countries (non-signatories) play as singletons. The equilibrium coalition size is found by applying the concept of internal&external stability.

---

19 The term is taken from Bloch (1997).

Definition 9: Stability in the Cartel Formation Game

Denote the non-trivial equilibrium coalition of signatories by c_S *,* $i \in c_S$ *, non-signatories by* $j \notin c_S$ *and let the coalition structure be given by* $c = (c_S, 1, ..., 1)$*.*

1) *Internal Stability: There is no incentive for a signatory to leave the coalition. That is,* $\pi_i(c_S, c) - \pi_i(1, c') \geq 0 \;\; \forall i \in c_S$ *where* $c' = c \setminus c_S \cup \{c_S \setminus \{i\}\} \cup \{i\}$
2) *External Stability: There is no incentive for a non-signatory to join the coalition. That is,* $\pi_j(c_S \cup \{j\}, c'') - \pi_j(1, c) < 0 \;\forall j \notin c_S$ *where* $c'' = c \setminus \{\{j\}, \{c_S\}\} \cup \{c_S \cup \{j\}\}$*.*[20]

From Definition 9 it is evident that internal stability corresponds to the definition of stand-alone stability (Definition 5). The definition of external stability implies de facto an open membership rule. That is, non-signatories may accede to the coalition if this is beneficial for them. From Definition 9 it is also apparent that only single deviations are considered when determining an equilibrium. That is, an internally and externally stable coalition structure is de facto a NE. Due to the simple structure of the game, it is straightforward to state the following result.

Proposition 8: Equilibrium Coalition Structure in the Cartel Formation Game

Let c_S^* *be the largest coalition for which* $c^* = (c_S^*, 1, ..., 1)$ *is stand-alone stable. Then* c^* *is the most concentrated equilibrium of the cartel formation game. Under condition* C_5 *the equilibrium is unique if* $c = (c_S, 1, ..., 1)$ *is stand-alone stable for all* $c_S < c_S^*$*.*

For payoff function [2] $c_S^* = 3$ *and for payoff function [3]* $c_S^* \in \{1, 2\}$*.*

Proof: c^* satisfies Definition 9.1 because it is stand-alone stable and also Definition 9.2 because $(c_S^* + 1, 1, ..., 1)$ is not stand-alone stable. Assume that there is an other coalition structure $c = (c_S, 1, ..., 1)$, $c_S \neq c_S^*$ that satisfies Definition 9. From Definition. 9.1 it follows that $c_S < c_S^*$. Hence c^* is the most concentrated equilibrium. If $c = (c_1, 1, ..., 1)$ is stand-alone stable for all $c_1 < c_S^*$, then c cannot be an equilibrium since singletons would join the coalition due to Assumption 2 and condition C_5 until the coalition is of size c_S^*. The final statement is an implication of the results above and Proposition 5 (**Q.E.D.**).

The result above is not terribly interesting since the equilibrium number of signatories is independent of the number of countries. Thus, one may wonder whether the result changes if some of the assumptions are modified. One possibility could be to assume an exclusive instead of an open membership rule. Exclusivity would imply that only if signatories are

---

[20] We use "<" instead "≤" to be consistent with Assumption 2.

willing to accept a non-signatory, an "external" player is allowed to join their club. However, it is easily checked that this modification has no effect in our context (under the conditions of Proposition 8).

A second possibility could be to assume a Stackelberg instead of a Nash-Cournot strategy of signatories in the second stage. Such an assumption has been made by Barrett (1994b, 1997b): signatories maximize the joint payoff of the coalition, taking the reaction of non-signatories into consideration. For payoff function [2] this change has no effect since reaction functions are orthogonal. For payoff function [3] one finds $c_s^* \in [2, N]$.[21] That is, the exact number of signatories depends on the parameter values N (total number of countries), b and c (benefit and cost parameter). The advantage of Barrett´s model version is that it can also explain IEAs which comprise more than 3 countries and that the coalition size can be related to the parameters of the model. It therefore allows drawing some conclusions of political relevance. For example a major finding derived from this version is that whenever cooperation would be needed most from a global perspective, a coalition achieves only little. The disadvantages of this version are: 1) For ex-ante symmetric countries it is difficult to justify why some countries (signatories) have more information than others (non-signatories). 2) The Stackelberg assumption implies irrational behavior of countries. For instance, consider the condition of internal stability. As a signatory, a country has an informational advantage as a Stackelberg leader but assumes that, provided it would become a non-signatory, it looses this information.[22]

A third possibility of a different assumption is derived by noting, as pointed out above, that Definition 9 implies a *myopic behavior* of players when deciding on their membership. Players only consider the immediate reaction to their change of membership but ignore possible chain reactions that may be triggered by their decision. For example assume N=5 and payoff function [2]. According to Proposition 8 $c_S^* = 3$. $c_S^* \neq 5$ since $\pi_i(5, (5)) < \pi_i(1, (4,1))$ and $\pi_j(4, (4, 1)) < \pi_j(1, (3, 1, 1))$. That is, viewing coalition formation as a sequential process starting from the grand coalition, players leave the coalition as long as a coalition is not stand-alone stable. However, a player of the grand coalition who does not only look one step ahead should realize that $\pi_i(5, (5)) > \pi_i(1, (3,1,1))$ and therefore may refrain from taking a free-ride.

---

21 For a comparison of equilibrium coalition structures for different payoff functions under the Nash-Cournot and Stackelberg assumption see Barrett (1997a) and Finus (2000b, ch. 13).

22 For an extensive discussion of this and related issues see Finus (2000b, ch. 13).

This feature of farsightedness has been proposed by Carraro/Moriconi (1997). We skip to give a formal definition of this modification since such a coalition formation game constitutes a special case of the more general case of an equilibrium binding agreement game (EBAG), which we discuss in subsection 3.3.2.[23] The only difference is that here the equilibrium number of non-trivial coalitions is restricted to one, whereas in the EBAG multiple coalitions are possible. In the present context, it may only be worthwhile to point out that the equilibrium coalition structures are determined recursively. That is, one starts by checking whether $c_S = 1$ is stable, which it is by definition. Thus, the interim largest stable coalition c_S^* is defined as $c_S^* = 1$ Then one checks for $c_S = 2$ by computing $F := \pi_i(c_S, c = (c_S, 1, ..., 1)) - \pi_i(1, c = (c_S^*, 1, ..., 1))$. If F≥0, then $c_S^* := c_S = 2$, otherwise we still have $c_S^* = 1$. For payoff function [2], F≥0, and therefore $c_S^* = c_S = 2$. Also for $c_S = 3$, F≥0, and hence $c_S^* = c_S = 3$. For $c^S = 4$, F<0 and thus $c_S^* = 3$. However, as argued above, for $c_S = 5$ and $c_S^* = 3$, F≥0 and therefore $c_S^* = 5$. More generally, we have:

Proposition 9: Equilibrium Binding Agreement in the Cartel Formation Game (Payoff Function [2]

a) In the cartel formation game an equilibrium binding agreement is given by $c^* = (c_S^*, 1, ..., 1)$ *where* $c_S^* \le N$ *is determined as follows:*

1) i:=0, $c_{S(i)}^* := 1$, $c_S := 1$.

2) $c_S := c_S + 1$

3) Let $F := \pi_i(c_S, c = (c_S, 1, ..., 1)) - \pi_i(1, c = (c_{S(i)}^*, 1, ..., 1))$. *If F≥0, then* $i := i + 1$ *and* $c_{S(i)}^* = c_S$. *As long as* $c_S < N$, *go to step 2. If* $c_S = N$ *stop.*

Then the most concentrated equilibrium coalition structure is given by $c^* = (c_S^*, 1, ..., 1)$ *where* $c_S^* = max\, c_{S(i)}^* \le N$**.**

b) For payoff function [2], F≥(<)0 if $c_S \ge (<) \left[1 + \sqrt{2{c_{S(i)}^*}^2 - 2c_{S(i)}^*} \right]$ *with [ ] the next lower integer (e.g., [3.5]=3 and [3]=3).*

c) The number of coalitions in c^* *is* $M^* \le I(N/3)$ *with I( ) the next higher integer (e.g., I(3.5)=I(4)=4 and I(3)=3)).*

Proof: a) Obvious and therefore omitted. b) Follows from the evaluation of F. c) is proved in Appendix 4 (**Q.E.D.**).

---

23 Though this modification should be grouped under sequential games, we discuss it here since the forces are best understood in connection with the ordinary cartel formation game.

For payoff function [2] we have $c^*_{S(i)}$={1, 2, 3, 5, 8, 12, 18, 26, 38, 55, …}. Thus for instance for N=5, $c^* \in$ {(1, 1, 1, 1, 1), (2, 1, 1, 1), (3, 1, 1), (5)} of which the most concentrated coalition structure is the grand coalition.

3.2.3 Open Membership Game

In the open membership game of Yi/Shin (1995) players can freely form coalitions as long as no outsider is excluded from joining a coalition. Players choose their membership by simultaneously announcing a message m_i (or in the diction of Yi/Shin they "announce an address"). Players that have announced the same message form a coalition. That is, if and only if $m_i=m_j$, then $\{i\} \cup \{j\} \subset c_k$. For instance, if N=4 and $m_1=m_2=m_3=1$ and $m_4=2$, c={{1, 2, 3}, {4}} forms. If country 3 changes its message to $m_3=2$, then c={{1, 2}, {3, 4}}.

Since a country can always leave a coalition by announcing a singleton address, a basic prerequisite for a coalition structure to qualify as a NE in the open membership game is that a coalition structure is stand-alone stable. For payoff function [2] this implies that only coalitions structures $c=(c_1, ..., c_M)$ with $3 \geq c_1 \geq ... \geq c_M$ qualify as NEs. For instance, suppose N=4, then c^1=(3, 1), c^2=(2, 2), c^3=(2, 1, 1) and c^4=(1, 1, 1, 1) are *potential* NEs. However, coalition structure c^1 cannot be a NE since a country belonging to the coalition comprising three countries has an incentive to announce the same address as (to join) the singleton country that follows from condition C_4. Though the singleton prefers to remain a singleton (which follows from conditions C_3), it cannot deny the accession under the open membership rule. Coalition structures c^3 and c^4 cannot be NEs since, given the announcements of the other countries, a singleton has an incentive to announce the same address as an other singleton due to C_4. Coalition structure c^2 is a NE since no country in a coalition comprising two countries has an incentive to become a singleton or to be a member of a coalition comprising three countries due to C_4. Due to $C^{CPNE} \subset C^{NE}$, coalition structure c^2 is the only candidate to qualify as a potential CPNE. It is easily checked that the only coalition structure that could challenge c^2 is the grand coalition which requires that two or four countries deviate. However, since the grand coalition is subject to further deviations, which are not in the interest of any initial deviator, we have $C^{CPNE} = c^2 = (2, 2)$. Since $c^2 \notin C^{PO}(4)$, there is no SNE.

Due to the simple structure of the game, one can derive quite general results (Yi 1997 and Yi/Shin 2000).

Proposition 10: Equilibrium Coalition Structure in the Open Membership Game (General Payoff Function)

In the open membership game a coalition structure must be stand-alone stable to be a NE coalition structure.

Suppose that $c^* = (c_1^*, ..., c_{M_1^*}^*, c_{M_1^*+1}^*, ..., c_{M^*}^*)$ *is a stand-alone stable coalition structure with* $c_1^* = ... = c_{M_1^*}^*$ *and* $c_{M_1^*+1}^* = ... = c_{M^*}^* = c_1^* - 1$. *Further suppose that each coalition structure* $c = (c_1, ..., c_{M_1}, c_{M_1+1}, ..., c_M)$ *with* $M < M^*$, $c_1 = ... = c_{M_1}$ *and* $c_{M_1+1} = ... = c_M = c_1 - 1$ *is not stand-alone stable, then*

a) *there exists a NE coalition structure under condition* C_4.

b) c^* *is the most concentrated NE coalition structure under condition* C_4 *and*

c) c^* *is the unique CPNE coalition structure under conditions* C_1 *to* C_4.

Proof: See Yi (1997) and Yi/Shin (2000).

The intuition of Proposition 10 is easy to grasp. c^* implies a rather symmetric coalition structure. This is due to two facts. First, a coalition structure must be stand-alone stable. Thus, if the size of the largest stand-alone stable coalition structure is substantially smaller than N, there will be several coalitions. Second, due to condition C_4, $c_1 \le c_M + 1$ ($c=(c_1, ..., c_M)$, $c_1 \ge ... \ge c_M$) can never be an equilibrium since a country of a bigger coalition would have an incentive to join a smaller coalition. By construction, c^* is the most concentrated stand-alone stable coalition structure satisfying $c_1 \le c_M + 1$. Now consider single deviations (Part b of Proposition 10). i) A country becomes a singleton which is not beneficial by the stand-alone property. ii) A member of one of the $c_1^*, ..., c_{M_1^*}^*$ coalitions joins one of the $c_{M_1^*+1}^*, ..., c_{M^*}^*$ coalitions which by symmetry does not change her payoff. iii) A member of one of the $c_{M_1^*+1}^*, ..., c_{M^*}^*$ coalition joins one of the $c_1^*, ..., c_{M_1^*}^*$ coalitions but then $c_1 \le c_M + 1$ is violated.
Since the singleton coalition structure c=(1, ..., 1) is stand-alone stable by definition, there will be at least one NE (Part a of Proposition 10).
The intuition of Part c of Proposition 10 is more difficult to crasp. In the case of deviations, which result in more concentrated coalition structures, it is evident that this is not feasible since the stand-alone stability condition would be violated. In the case of deviations which result in less concentrated coalition structures it can be shown that this implies a welfare loss to players involved in the deviation.

In the context of payoff function [2], Proposition 10 reads as follows.

Proposition 11: Equilibrium Coalition Structure in Open Membership Game (Payoff Function [2])

a) The largest equilibrium coalition structure comprises no more than 3 countries. b) A coalition structure with two (or more) singletons can never be an equilibrium coalition structure. c) A coalition structure with a coalition of size one and a coalition of size two can be a NE but never a CPNE. d) Let k=[N/3] and $R = N - k \cdot 3$ and assume N≥3, then the unique CPNE coalition structure is given by i) R=0: $C^* = (c_k)$ with $c_k = 3$, ii) R=1: $C^* = (c_{k-1}, 2, 2)$ with $c_{k-1} = 3$ and R=2: $C^* = (c_k, 2)$ with $c_k = 3$. For N=2, $C^* = (2)$. e) For N>3 there is no Pareto-optimal CPNE and no SNE coalition structure.

Proof: a) Follows from Propositions 10 and 5. b) Follows from condition C_4. c) Such a coalition structure can be a Nash equilibrium since there is no incentive for a single deviation (e.g., N=5, c={{1, 2}, {3, 4}, {5}}). However, a coalition of two members always gains if the members announce the same address as the singleton according to C_5. d) Follows from Proposition 10. e) The first part of the statement follows from d) and the necessary condition of a PO for payoff function [2]: $c_i > 2c_{i+1}$, for any i=1...M-1, where $c=(c_1, ..., c_M)$. The second part of the statement follows from the first part of the statement and $C^{SNE} \subset C^{CPNE}$ (**Q.E.D.**).

A list with the equilibrium coalition structures for $N \in \{2, ..., 12\}$ is provided in Table 3, section 4.

3.2.4 Exclusive Membership Δ-Game

In the exclusive membership Δ-game of Hart/Kurz (1983) players simultaneously announce a list of coalition members with whom they like to form a coalition, $\ell_i \subset N$. Those players who have each other on the list will form a coalition. That is, if and only if $j \in l_i$ and $i \in l_j$, then $\{i, j\} \subset c_k$. For instance suppose N=4 and $\ell_1 = \{1, 2, 3\}$, $\ell_2 = \{1, 2, 3\}$, $\ell_3 = \{3\}$ and $\ell_4 = \{3, 4\}$, then $c = \{\{1, 2\}, \{3\}, \{4\}\}$ forms. Players 1 and 2 have both each other on the list and therefore form a coalition. Players 3 and 4 remain singletons. Though player 4 and also players 1 and 2 would like to form a coalition with player 3, player 3 can remain a singleton since membership is exclusive. In other words, *a coalition only forms by unanimious agreement*. Thus in contrast to the open membership game, a country can only join an other coalition if all members agree. For instance, consider our previous example N=4 and payoff function [2]. In the open membership game coalition structure $c^1 = (3, 1)$ was not a NE since a member of the bigger coalition likes to join the singleton coalition by C_4, and the singleton cannot deny access. In the exclusive membership Δ-game, coalition structure $c^1 = (3, 1)$ is also a NE since the singleton will deny access to his coalition because of $\pi_i(1, (3, 1)) > \pi_i(2, (2, 2))$ due to C_3.

Like in the open membership game, in the exclusive membership Δ-game a basic prerequisite for a stable coalition structure is that a coalition structure is stand-alone stable: a player can always announce a list including only himself which is beneficial for him if a coalition structure is not stand-alone stable. Moreover, by the construction of the exclusive membership Δ-game for every stand-alone stable coalition structure c there exists a set of announcements that supports c as a Nash equilibrium outcome. Suppose each player announces a list with only those players on it which actually form the coalition. Then no player can join an other coalition by an individual deviation. The only feasible deviation is to form a singleton coalition which is not beneficial if a coalition structure is stand-alone stable. Since the set of stand-alone stable coalition structures is usually large, the equilibrium refinements CPNE and SNE are particularly useful in this coalition game.

For instance, suppose N=4 and payoff function [2], then c^1=(3, 1), c^2=(2, 2), c^3=(2, 1, 1) and c^4=(1, 1, 1, 1) are stand-alone stable coalition structures and hence NEs. However, c^3 and c^4 are not CPNEs since singletons have an incentive to merge by C_4. c^1 is a CPNE since the singleton has no incentive join the larger coaltion by C_4 and the members of the coalition comprising three countries can neither enforce the grand coalition nor have they an incentive to form smaller coalitions by C_5. Similar arguments establish that c^2 is also a CPNE. Since c^1 is the only stand-alone PO and, as shown above, no subgroup of players has an incentive to jointly deviate, c^1 is the only SNE. More generally, we find (Yi/Shin 2000):

Proposition 12: Equilibrium Coalition Structure in the Exclusive Membership Δ-Game (General Payoff Function)

Under the exclusive membership Δ-rule the set of NEs is non-empty and equal to the set of stand-alone stable coalition structures.

Suppose that $c^* = (c_1^*, ..., c_{M-1}^*, c_{M^*}^*)$ *is a stand-alone stable coalition structure with* $c_1^* = ... = c_{M-1}^* \geq c_{M^*}^*$. *Further suppose that any coalition structure c for which* $c_1 > c_1^*$ *is not stand-alone stable, then*

a) c^* *is the most concentrated NE.*

b) c^* *is the most concentraded CPNE under conditions* C_1 *to* C_3*.*

c) If there is any other CPNE it comprises exactly M^* *coalitions and is less concentrated than* c^* *under conditions* C_1 *to* C_3*.*

d) If $c_1^* - 1 \leq c_{M^*}^* \leq c_1^*$, *then* c^* *is the unique coalition-proof Nash equilibrium coalition structure under conditions* C_1 *to* C_3*.*

Proof: See Yi/Shin (2000).

The intuition behind the proof of Proposition 12 is the following. The second part of the general remark in Proposition 12 has already been explained above. The first part simply follows from the fact that the singleton coalition structure is stand-alone stable by definition. Therefore the existence of a NE coalition structure is guaranteed. Statement a) follows from two facts. First, coalition structure c^* is stand-alone stable by assumption and therefore a NE. Second, by the construction of c^*, this is the most concentrated coalition structure with M^* coalitions and the largest coalition being of size c_1^*. Consequently, any other coalition structure will either be less concentrated or will include a coalition of size larger than c_1^* which by assumption is not stand-alone stable. c^* is also a CPNE (statement b, part 1) since any deviation which creates a bigger coalition is not stand-alone stable and any coalition which creates less concentrated coalition structures is not profitable by Assumption 2 and conditions C_1 to C_3. Since $C^{CPNE} \subset C^{NE}$, c^* is also the most concentrated CPNE (statement b, part 2). Statement c) follows from two facts: First, if there was an other CPNE coalition structure c, c cannot comprise less than M^* coalitions since then stand-alone stability would be violated. Second, c cannot comprise more than M^* coalitions because each country could benefit from a joint deviation to c^* if the members of the smallest coalitions in c join the smallest coalition in c^* and so on. Statement d) follows from the fact that if $c_1^* - 1 \le c_{M^*}^* \le c_1^*$, then c^* is the unique coalition structure with M^* coalitions and with the size of the largest coalition equal to c_1^*.

In the context of payoff function [2] Proposition 12 reads as follows.

Proposition 13: Equilibrium Coalition Structure in the Exclusive Membership Δ-Game (Payoff Function [2])

a) The largest coalition of an equilibrium coalition structure comprises no more than 3 countries. b) A coalition structure with two singletons or one singleton and a coalition of size two can never be a CPNE. c) A coalition structure comprising of more than two coalitions of size two can be a NE but not a CPNE. d) Let k=[N/3], $R = N - k \cdot 3$*, then the equilibrium coalition structure is given by R=0:* $c^* = (c_k) = (3, 3, ..., 3)$ *(k coalitions of size 3), R=1:* $c^* \in \{(c_k, 1), (c_{k-1}, 2)\}$ *and R=2:* $c^* = (c_k, 2)$*. For N=2,* $c^* = (2)$*. e) For N>4 there is no Pareto-optimal CPNE and no SNE.*

Proof: a) Follows from Propositions 5 and 12. b) Follows from C_4 (and Assumption 2). c) A coalition structure of three or more coalitions of size two can be a Nash equilibrium. For example for N=6 it is easily checked that c=(2, 2, 2) is a Nash equilibrium coalition structure. However, a coalition structure c=(2, 2, 2, c′) cannot be a CPNE since it is Pareto-dominated by c=(3, 3, c′). d) Follows from Proposition 12. e) A similar argument as in the case of the open-membership game applies (**Q.E.D.**).

A list with the equilibrium coalition structures for N∈{2, ..., 12} is provided in Table 3, section 4.

3.2.5 Exclusive Membership Γ-Game

The exclusive membership Γ-game goes back to Von Neumann/Morgenstern (1944) and has been reintroduced by Hart/Kurz (1983) under this name. This game is very similar to the exclusive membership Δ-game. That is, each player simultaneously announces a list of coalition members with whom she likes to form a coalition, $\ell_i \subset N$. The only difference is that a coalition only forms if and only if there is unanimious agreement among all prospective members of a coalition to form exactly this coalition. For instance suppose our previous example N=4 and $\ell_1 = \{1,2,3\}$, $\ell_2 = \{1,2,3\}$, $\ell_3 = \{3\}$ and $\ell_4 = \{3,4\}$, then, in contrast to the exclusive membership Δ-game where $c = \{\{1,2\},\{3\},\{4\}\}$ formed, $c = \{\{1\},\{2\},\{3\},\{4\}\}$ in the exclusive membership Γ-game. This implies that if a subgroup or subgroups of players jointly deviate, then the coalitions to which they belonged will break apart. Because of this strong assumption, the amount of NE is usually very large and very concentrated coalition structures can be supported as equilibria.

For instance, consider our previous example of N=4 and payoff function [2]. All coalition structures, c^1=(3, 1), c^2=(2, 2), c^3=(2, 1, 1), c^4=(1, 1, 1, 1) and c^5=(4) are NE. The grand coalition is a NE since $\pi_i(4) > \pi_i(1,(1,1,1,1))$. For the other coalition structures, it is easily checked that also no player has an incentive to leave his coalition. The grand coalition is also a SNE since a deviation by a subgroup of players does not pay. The resulting coalition structure would be less concentrated, the deviators would be members of the largest coalition and therefore would have an incentive to remerge by condition C_5. c^1=(3, 1) is also a SNE since the singleton has no incentive to form a grand coalition or to form a coalition with any other subgroup of players by conditions C_1 and C_4. For the members of the coalition comprising three countries the same arguments apply as in the case of the grand coalition. Coalition structures c^2, c^3 and c^4 cannot be SNEs since they are strictly Pareto-dominated by the grand coalition (due to C_5). From $C^{SNE} \subset C^{CPNE}$, it follows that c^1 and c^5 are also CPNE. Since c^2, c^3 and c^4 are strictly Pareto-dominated by the grand coalition, which itself is a CPNE, no other coalition structure except c^1 and c^5 can be a CPNE.

Proposition 14: Equilibrium Coalition Structure in the Exclusive Membership Γ-Game (General Payoff Function)

a) There are at least two NE of which the grand coalition is the most concentrated one.

b) The grand coalition is a CPNE and a SNE under condition C_1.

c) All coalition structures are a NE under condition C_5.

d) The set of CPNE is equal to the set of SNE coalition structures and comprises the set of Pareto-optimal coalition structures under condition C_1 and C_5.

Proof: a) In the exclusive membership Γ-game a coalition only forms if all coalition members announce the same list of coalition members. If one member annouces a different list, this implies de facto that the coalition breaks apart. Thus, the singleton coalition structure is a Nash equilibrium by definition. Moreover, since $\pi_i(n,(N)) > \pi_i(1,(1,...,1))$ holds for any payoff function in the symmetric global emission game, the grand coalition constitutes a Nash equilibrium in the the exclusive membership Γ-game. b) A deviation by a group of players, say n, from the grand coalition leads to a coalition structure $c = (c_1,...,c_M)$ with $c_1 \geq ... \geq c_M$ and $c_{M-(N-n)+1} = ... = c_M = 1$ if $n < N$. Since $N\pi_i(N,(N)) > c_1\pi_i(c_1,c) + ... + c_M\pi_j(c_M,c)$ and $\pi_i(c_1,c) \leq \pi_j(c_j,c)\ \forall\ j = 2,...,M$ by C_1, the deviators who form c_1 are worse off. Since the grand coalition is a PO by Proposition 6, the grand coalition is a SNE and also a CPNE (due to $C^{SNE} \subset C^{CPNE}$). c) A Nash equilibrium coalition must satisfy $\pi_i(c_i,c) \geq \pi_i(1,c')\ \forall\ c_i$ with $c' = c \setminus c_i \cup \{1,...,1\}$ which holds under condition C_5. d) Assume that there exists a coalition structure $c \in PO(C^{NE})$ which is not a SNE. Then there exists a set of deviators n which can improve the payoff of at least one member without worsening the payoff of other members by deviation to the coalition structure c'. Coalition structure c and c' may be written as $c = (c^n, \tilde{c})$ and $c' = (c'^n, \tilde{c})$ with c^n the set of coalitions which have at least one deviating member. c'^n comprises the coalitions of the deviators and singleton coalitions. Since the deviators receive at least the same payoff in c' as in c the same must be true for the singletons due to C_1. Hence the residuals $\tilde{c}$ must receive a lower payoff in c' since otherwise c would be Pareto-dominated by c' (violating our initial assumption). From this it follows that global emissions are higher in c' than in c. Consequently, emissions of $\tilde{c}$ are lower in c' than in c. In order to benefit from the move from c to c' the deviators must increase their emissions since global emissions are higher in c than in c'. Thus, $c_i' \leq c_i\ \forall\ i \in N$ because of condition C_1. Hence, a deviation including members of several coalitions is less profitable than a deviation of members of a single coalition. Hence we may assume $c^n = c_i$. Then the deviation is not profitable at least for the largest coalition of c'^n due to C_5. (**Q.E.D.**).

For payoff function [2] we find:

Proposition 15: Equilibrium Coalition Structure in the Exclusive Membership Γ-Game (Payoff Function [2])

In the exclusive membership Γ-game the set of NE coalition structures comprises all coalition structures. The set of CPNE and SNE coalition structures comprises the set of Pareto optimal coalition structures as given in Proposition 7.

Proof: From Proposition 2 it follows that payoff function satisfies conditions C_1 to C_5 and thus Proposition 14 can immediately be applied (**Q.E.D.**).

A list with the equilibrium coalition structures for $N \in \{2, ..., 12\}$ is provided in Table 3, section 4.

3.3 Sequential Choice of Membership

3.3.1 Sequential Move Unanimity Game

The *sequential move unanimity game* (SMUG) goes back to Bloch (1995, 1996). The game is in spirit of Rubinstein´s (1982) two-player alternating-offers bargaining game and is a generalization of Chatterjee et al.´s (1993) extension to an N-country bargaining game. The game proceeds as follows. First, countries are ordered according to some (external) rule, e.g., countries are indexed. The country with the lowest index (initiator) starts by proposing a coalition to which she wants to belong. Each prospective member is asked whether it accepts the proposal. According to the external rule the country with the lowest index in the prospective coalition is asked first, then that with the second lowest index and so forth. If all prospective members agree, the coalition, say c_i, is formed and the remaining players $N \setminus c_i$ may form coalitions among themselves. The country with the lowest index among $N \setminus c_i$ becomes the new initiator. If a country rejects a proposal, it can make a new proposal. That is for a coalition to form, unanimous agreement is required which corresponds to the assumption in the exclusive membership Γ-game. A sequence of proposals finally leads to a coalition structure if the proposals form a subgame-perfect equilibrium.[24]

Basically, if an initiator makes a proposal she will think about two things. First, will her proposal be acceptable to the potential members? Obviously, a proposal which is unacceptable will only imply that the initiative to make a proposal is passed on to the next player which is not in the interest of the proposer. Second, suppose the players who are asked for acceptance agree with the proposal. Then the question arises which coalition the remaining players will form? The answer to this question will of course affect the proposal at the

---

[24] For a formal definition see Bloch (1996), Finus (2000b, ch. 15) and Ray/Vohra (1999).

initial stage. Thus, an initiator must solve the entire game backward for all players to find his best strategy.

For ex-ante symmetric players Bloch (1996) has shown that a simple finite procedure can be used to determine the equilibrium/equilibria in the infinite SMUG. The algorithm works as follows. The first player proposes the size of his coalition $c_1 \in [1, N]$. Then player c_1+1 proposes a coalition of size $c_2 \in [1,\ N-c_1]$. This process continues until $c_1+c_2+...+c_M$ =N.

The intuition why the infinite SMUG reduces to a simple "size announcement game" is the following. Due to symmetry 1) payoffs to a player depend only on the sizes of coalitions and not on individual members, 2) the interests of the proposer and those countries which are asked for acceptance coincide and 3) each player prefers to be asked first since the proposer makes an offer which serves his interests best. In the global emission game countries prefer to be a member of the smallest coalition for a given equilibrium coalition structure (see condition C_1). Points 2 and 3 are responsible for no delay equilibria: players make only proposals which are immediately accepted.

For instance, consider payoff function [2] and N=4. First note that $\pi_i(3,(3,1)) = \pi_i(1,(2,1,1)) > \pi_i(2,(2,1,1)) > \pi_i(1,(1,1,1,1))$ holds (due to conditions C_1 and C_4). Second note that $\pi_i(4,(4)) < \pi_i(1,(3,1))$ is true (due to Proposition 5). Consequently, if the first initiator proposes to remain a singleton, the remaining countries $c\backslash\{1\}$ will either form $c\backslash\{1\} = (3)$ or $c\backslash\{1\} = (2,1)$. However, due to Assumption 2, we select $c\backslash\{1\} = (3)$ and $c = (3,1)$ is the unique equilibrium coalition structure.

Unfortunately, not very much can be said about the equilibrium coalition structure in the SMUG at a general level. Therefore, Bloch (1997) characterizes the equilibrium structures for a public goods model which exhibits a similar structure than our global emission game assuming payoff function [2]. In both models reaction functions are orthogonal and a coalition chooses its optimal strategy, that is, its contribution to the public good and its emission level, respectively, independent of the strategy of the other coalitions. Bloch suggests that for Assumption 2 the equilibrium coalition structure is given by the Fibonnaci decomposition which is derived from a sequence of Fibonnaci numbers where $f_0 = 1$, $f_1 = 2$ and $f_i = f_{i-1} + f_{i-2}$. Thus, the Fibonnaci numbers are given by 1, 2, 3, 5, 8, 13, 21 and so on. According to Bloch, the equilibrium coalition structure is derived as follows. One starts by choosing the largest Fibonnaci number equal or smaller than N which is denoted by f_F^1. Then one looks for the largest Fibonnaci number equal or smaller than $N - f_F^1$. This process continues until $\Sigma f_F^i = N$.

For instance, for $N = 4$, the largest Fibonnaci number is 3 and $N - 3 = 1$. Hence, the equilibrium coalition structure would be $c = (3, 1)$. For $N = 20$ Bloch predicts $c = (13, 5, 2)$. However, computations reveal that in Bloch´s setting as well as in our setting $\pi_i(2, (13, 5, 2)) < \pi_i((20))$ holds. That is, a member of the smallest coalition in a coalition structure $c = (13, 5, 2)$ strictly prefers the grand coalition and by C_1 all other players as well. Consequently, $c = (13, 5, 2)$ cannot be an equilibrium coalition structure since the initiator would propose the grand coalition instead of a coalition comprising himself and one other country and this proposal would be accepted. This leads to the following proposition.

Proposition 16: Equilibrium Coalition Structure in the Sequential Move Unanimity Game (General Payoff Function)

In the sequentially move unanimity game (SMUG), the equilibrium coalition structure must be a Pareto-optimal coalition structure.

Proof: Assume a coalition structure c which is an equilibrium in the SMUG but not a PO. Then there exists an other coalition structure c' with $\pi_i(c_i, c) \leq \pi_i(c_i', c') \ \forall i$ and $\pi_j(c_j, c) < \pi_j(c_j', c')$ for at least one player j. Hence if the initiator proposes c', no country has an incentive to reject the proposal and instead proposing the smallest coalition of coalition structure c. Consequently, c cannot be a SMUG due to Assumption 2 (**Q.E.D.**).[25]

Interestingly, for payoff function [2] and N<20 the equilibrium coalition structure coincides with the Fibonnaci decomposition under Assumption 2. That is, this decomposition belongs to the set of POs. However, for N≥20 this is not true any longer. Therefore, we invent a decomposition which we call Pareto dominance (PD) decomposition and which is derived from a sequence of PD-numbers to characterize the equilibrium coalition structure for payoff function [2]. We denote the PD-numbers by $f = \{f_0, ..., f_i\}$ in ascending sequence, e.g., $f = \{1, 2, 3, 5, 8, 13, 20, 31, ...\}$ and the PD-decomposition by $\Phi(N) = \{f^1(N), ..., f^M(N)\} = \{f^1, ..., f^M\}$ in descending sequence. For $N \in \{2, ..., 12\}$ equilibrium coalition structures are summarized in Table 3. More generally, we have:

[25] Remark: Dropping Assumption 2, then an equilibrium in the SMUG must not be strongly Pareto-dominated by any other coalition structure. That is, there is no coalition structure c' with $\pi_j(c_j, c) < \pi_j(c_j', c') \ \forall j$.

Proposition 17: Equilibrium Coalition Structure in the Sequential Move Unanimity Game (Payoff Function [2])

a) In the sequential move unanimity game the unique equilibrium coalition structure is given by $c^* = \Phi(N)$ *where* $\Phi(N)$ *is derived as follows. Fix* $i := 0$ *and define* $f_i := 1$ *and* $\Phi(1) := 1$. *Let the Pareto dominance decomposition of N be given by* $\Phi(N) = \{f^1(N), ..., f^M(N)\}$, $f^k > f^{k+1}$ $\forall$ $k<M$, $\{f^1, ..., f^M\} \subset \{f_0, ..., f_i\}$ *where* $f^1 = max\, f_j \le N$, $f^2 = max\, f_j \le N - f^1$ *etc. Then*

$$\Phi(N+1) = \begin{cases} \{f^1(N), \Phi(N+1-f^1(N)\} & \text{if this coalition structure is not weakly Pareto-} \\ & \text{dominated by the grand coalition} \\ \{N+1\} & \text{otherwise} \end{cases}$$

If $\Phi(N+1) = \{N+1\}$, *then* $i := i+1$ *and* $f_i := N+1$.

b) Let $c^* = (c_1, ..., c_M)$, *then* $c_i \ge \frac{9}{4} c_{i+1}$.

c) The equilibrium coalition structure comprises no more than one coalition of size equal or smaller than two.

Proof: See Appendix 5 (**Q.E.D.**).

For practical purposes if for a given N the sequence of PD-numbers is already known, the PD-decomposition can simply be derived by chosing the largest PD-number smaller or equal to N, i.e., $f^1 \le N$. Then one searches for the largest PD-number equal of smaller than N-f^1(N) and so on. This process continues until $\Sigma f^i = N$. For payoff function [2], the PD-numbers, $\{f_0, ..., f_i\}$, are given by {1, 2, 3, 5, 8, 13, 20, 31, 47, 73, ...} if damages are sufficiently large.[26] From this it follows that the PD-decomposition for N=53 is $c^* = \Phi(53) = \{47, 5, 1\}$ and for N=64 $c^* = \Phi(64) = \{47, 13, 3, 1\}$.

Intuitively, the equilibrium coalition structure in the SMUG may be explained as follows. Each player likes to be a member of the smallest coalition due to C_1 and prefers if the remaining players form a concentrated coalition structure due to C_2. Thus, a proposer has an

---

[26] For payoff function [2] and $f_i > f_6 = 20$, PD-numbers depend on the parameter values. For instance, for $f_7 = 31$ we have to require d>(37/60)c/b, for $f_8 = 47$ d>(105/86)c/b and for $f_9 = 73$ d>(85/144)c/b. That is, the larger damages are compared to the cost-benefit ratio the higher are the incentives for cooperation which results in smaller steps between the PD-numbers. Since this effect appears only for large coalitions it does not affect the comparison of equilibrium coalition structures among the sequential move unanimity game, open membership game, exclusive membership Δ-game and exclusive membership Γ-game.

incentive to propose a small coalition, however, subject to the constraint that remaining players form a stable coalition structure. A coalition structure is stable if and only if it comprises PD-numbers. Since according to Proposition 16, the equilibrium coalition structure must be a PO, only those POs which comprise exclusively PD-numbers are potential equilibrium coalition structures in the SMUG. Consequently, the proposer will choose that PO which allows it to be a member of the smallest coalition. For instance for N=12, $c = (12)$ and $c = (8, 3, 1)$ of which the proposer prefers the second coalition structure. In contrast, for N=8, $c = (8)$ is the only PO comprising exclusively PD-numbers and therefore the proposer will propose the grand coalition. Since for a given N the PD-decomposition is unique, there is only one equilibrium coalition structure.

A list with the equilibrium coalition structures for $N \in \{2, ..., 12\}$ is provided in Table 3, section 4.

3.3.2 Equilibrium Binding Agreement Game

Ray/Vohra (1997) motivate their *equilibrium binding agreement game* (EBAG) with the help of the following story. Initially the grand coalition gathers. Then some *leading perpetrators* may propose a different coalition structure if this is in their interest. The perpetrators split up to form a coalition, say c_i. In a next step, either members of the coalition c_i or the coalitions $C \setminus c_i$ may propose further deviations. Those countries which initiate further deviations are called *secondary perpetrators*. The process of disintegration continues until a coalition structure has been reached where no country likes to split up into finer partitions. Such a coalition structure constitutes an equilibrium binding agreement (EBA).[27]

Important for the understanding of the game is the assumption that coalitions can only become finer but not coarser and that only members of the same coalition can form smaller coalitions but this is not possible across coalitions. Similar to Carraro/Moriconi´s extension of the cartel formation game, countries are assumed to be farsighted in that they do not deviate from a given coalition structure if the final outcome (resulting from possible further

[27] At this stage it is already apparent that the term equilibrium binding agreement is a misnomer - it conveys the conjecture that the EBAG belongs to the realm of cooperative game theory. However, except for the weakness which is pertinent to all coalition games treated in this paper, namely that stability within the component game is either assumed ad hoc or via the assumption of instant reactions by players, the EBAG clearly belongs to non-cooperative game theory. An EBA equilibrium must be a self-enforcing coalition structure.

deviations) implies a payoff loss to them. Leading perpetrators and secondary perpetrators deviate if it is in their interest to do so. Thus, reactions of players are consistently defined.

More formally, like a CPNE coalition structure, equilibrium coalition strucures in an EBAG are recursively defined (Ray/Vohra 1997).[28]

Definition 10: Equilibrium Binding Agreement (EBA)

1) The finest coalition structure only consisting of singletons, $c^* = (1, ..., 1)$*, is an equilibrium binding agreement.*

2) Consider coalition structures c which have c^* *as their only refinement.* c^* *blocks c if there exists a perpetrator* c_i^* *such that* $\pi_i(c_i^*, c^*) > \pi_i(c_i, c) \quad \forall i \in c_i^*,\ c_i^* \subset c_i$

Recursively, suppose that for some coalition structure c the set C' *containing all refinements of c which are equilibrium binding agreements has been determined.*

Then c is blocked by $c' \in C'$ *if*

a) there is a leading perpetrator c_i' *who gains from the deviation, i.e.,* $\pi_i(c_i', c') > \pi_i(c_i, c)$ $\forall i \in c_i',\ c_i \subset c_i'$ *and*

b) any remerging of the secondary perpetrators is blocked by c' *as well. (Secondary perpetrators are those countries which are in* c' *members of smaller coalitions than in c. Formally, j is a secondary perpetrator if* $j \notin c_i'$, $j \in c_j' \subset c_j$ *and* $c_j' \neq c_j$.

c is an equilibrium binding agreement, if and only if there exists no refinement $c' \in C'$ *that blocks c.*

Ray/Vohra (1997), section 7, provide an algorithm do determine the EBAs. In the following we apply this algorithm to an example, assuming payoff function [2] and N=6, which is displayed in Table 2.

In the first column all coalition structures are displayed. The algorithm proceeds from the second column to the right. Define $\widehat{C}_1 = \{c^1\} = \{(1, ..., 1)\}$ where the tilt denotes EBAs. For C^2 select all coalitions which cannot be split up into any other coalition structure except $\widehat{C}^1$ and which are not blocked by $\widehat{C}^1$. In the example $C^2 = \{c^2, ..., c^{11}\}$. For $\widehat{C}^2$ select those coalitions of C^2 which cannot be split up into any other coalition of C^2. In the example $\widehat{C}^2 = \{c^2\}$. For C^3 select all coalitions of $C^2 \setminus \widehat{C}^2$ which are not blocked by $\widehat{C}^2$. In the example $C^3 = \{c^3, ..., c^{11}\}$. For $\widehat{C}^3$ select those coalitions of C^3 which cannot be split up into

[28] For a full account of the concept the reader is referred to Ray/Vohra (1997).

any other coalition of C^3. In the example $\widehat{C}^3 = \{c^3, c^4\}$. The process continues until C^k is empty. In Table 2 all EBAs $\widehat{C}(N) = \{\widehat{C}^1 \cup \widehat{C}^2 \cup ... \cup \widehat{C}^{k-1}\}$ are marked bold in the first column.

Table 2: Algorithm for Determining Equilibrium Binding Agreements*

	$\in \widehat{C}^1$	$\in C^2$	$\in \widehat{C}^2$	$\in C^3$	$\in \widehat{C}^3$	$\in C^4$	$\in \widehat{C}^4$	$\in C^5$	$\in \widehat{C}^5$	$\in C^6$
c^{11}=(6)		•		•		•		•		
c^{10}=(5, 1)		•		•		•		•	•	
c^9=(4, 2)		•		•		•				
c^8=(3, 3)		•		•		•		•	•	
c^7=(4, 1, 1)		•		•						
c^6=(3, 2, 1)		•		•		•	•			
c^5=(2, 2, 2)		•		•		•	•			
c^4=(3, 1, 1, 1)		•		•	•					
c^3=(2, 2, 1, 1)		•		•	•					
c^2=(2, 1, 1, 1, 1)		•	•							
c^1=(1, 1, 1, 1, 1 1)	•									

* Payoff function [2] and N=6 is assumed.

Note that $c^9 = (4, 2)$ is not an EBA since it is blocked by $c^6 = (3, 2, 1) \in \hat{C}_4$. However, $c^9 = (4, 2) \in C^4$ since it is not blocked by any coalition in $\widehat{C}^3$. Though $\pi_i(4, (4, 2)) < \pi_i(1, (3, 1, 1, 1))$, such a deviation would require that, after a perpetrator deviated from $c^9 = (4, 2)$ to $c^6 = (3, 2, 1)$, a secondary perpetrator would deviate from $c^6 = (3, 2, 1)$ to $c^4 = (3, 1, 1, 1)$. This is, however, not in the interest of the secondary perpetrator. With respect to Definition 10 the secondary perpetrator of the singleton coalition in $c^4 = (3, 1, 1, 1) \in \widehat{C}^3$ would have an incentive to remerge to c^6=(3, 2, 1).

Since the number of EBAs is usually quite large, it seems sensible to introduce a selection device. We therefore display only the coarsest coalition structures in Table 3 (section 4) which we denote $\tilde{C}^{(EBA)}(N)$. For the example $\tilde{C}^{(EBA)}(N = 6) = \{c^{10} = (5, 1), c^8 = (3, 3)\}$. This selection device is motivated by the algorithm used to determine EBAs. Instead of viewing the coalition formation process as starting from the grand coalition (see the motivation of EBAs at the beginning of this section), one could, alternatively, view this process as starting from the singleton coalition. Coalitions gradually become larger until no further enlargement is stable. Coalitions which merge will only do so if this is beneficial for them and outsiders benefit by condition C_1 (see section 2).

Despite this selection device, the prediction of equilibrium coalition structures is not as sharp as in the SMUG.

Proposition 18: Equilibrium Coalition Structure in the Equilibrium Binding Agreement Game (General Payoff Function)

a) *A stand-alone coalition structure is an equilibrium binding agreement under conditions C_1 to C_3.*

b) *For a given N, the set of coarsest equilibrium binding agreements, $\tilde{C}^{(EBA)}(N)$, may contain coalition structures which are no PO, i.e., $c \in \tilde{C}^{(EBA)}(N)$, $c \notin PO(N)$.*

c) *Those non POs may lead to a higher global welfare than a coalition stucture which constitutes a PD-decomposition, i.e., $c \in \tilde{C}^{(EBA)}(N)$, $c' = \Phi(N)$. That is $\Sigma\pi_i(c) > \Sigma\pi_i(c')$.*

Proof: For part a see Yi (1997). Parts b and c follow by example. Suppose payoff function [2]. Then for N=6 $\tilde{C}^{(EBA)}(N) = \{(5,1),(3,3)\}$ where (3, 3) is not a PO according to Table 1. For N=12, we find $\tilde{C}^{(EBA)}(N) = \{(8,3,1),(6,6)\}$ and $12\pi_i(6,(6,6)) > 8\pi_i(8,(8,3,1)$ $+3\pi_i(3,(8,3,1)) + \pi_i(1,(8,3,1))$ (**Q.E.D.**)

So far, we have not been able to characterize the exact nature of the coarsest EBAs. As for N∈ {2, ..., 12}, we also found for larger N that the set of coarsest EBAs always contains the SMUG equilibrium. Apart from the SMUG equilibrium, there may also be a not Pareto-efficient EBA belonging to the set of coarsest EBAs. We did not find any example were either 1) a PO, which is not a SMUG, or 2) only a non-PO but not a SMUG was among the set of coarsest EBAs. Thus, we suspect that it should be possible for payoff function [2] to prove that every SMUG is also a EBA and will belong to the set of coarsest EBAs and that if there is a non-PO belonging to the set of coarsest EBAs this will not be more concentrated than the SMUG. However, what is already evident from Proposition 18c is that there are EBAs which do not belong to the set of Pareto optimal coalition structures but generate a higher welfare.

4. Comparison, Discussion and Evaluation of Coalition Games

4.1 Comparison

As in previous sections, we group results under those which can be derived at a general and those which can be derived for payoff function [2]. For the discussion it is helpful to recall some results and assumptions of section 2. First, we assumed symmetric countries. Second, the coarser (the more concentrated) a coalition structure, the lower are global emissions for the general payoff function (payoff functions [2] and [3]) (Proposition 1). Third with respect to welfare such a relation could only be established for payoff function [2] (Proposition 3).

More precisely, the more concentrated a coalition structure is, the higher is global welfare assuming payoff function [2]. For payoff function [3] and other payoff functions it is easily checked that this relation does not generally hold. We only know that the grand coalition produces the highest welfare among all coalition structures (Proposition 4).

To ease the subsequent discussion, we display equilibrium coalitions for payoff function [2] and $N \in \{2, ..., N\}$ in Table 3 and introduce some notation.

Let $M^*(G(N))$ denote the set of most concentrated equilibria of a coalition formation game G for a given number of countries N. We write $c^1 \pi c^2$ ($c^1 \pi\pi c^2$), iff c^2 is more concentrated (coarser) than c^1, and $c^1 \approx c^2$ iff c^1 and c^2 cannot be compared under concentration. More specifically:

$M(g)\pi(\pi)M(h)$ (g and h are coalition formation games) iff $\forall$ N, $\forall\, c^g \in M^*(g(N))$, $\forall c^h \in M^*(h(N)): c^g \pi(\pi) c^h$ or $c^g = c^h$ and $\exists$ N, $c^g \in M^*(g(N))$, $c^h \in M^*(h(N)): c^g \pi(\pi) c^h$.

$M(g)\pi\phi M(h)$ iff $\exists\, N, \tilde{N}$, $c^1 \in M^*(g(N), c^2 \in M^*(h(N)$, $\tilde{c}^1 \in M^*(g(\tilde{N}), \tilde{c}^2 \in M^*(h(\tilde{N})$: $c^1 \pi c^2$ and $\tilde{c}^1 \phi \tilde{c}^2$ or $c^1 \approx c^2$.

Proposition 19: Comparison of the Most Concentrated Equilibria in Different Coalition Formation Games (General Payoff Function and Payoff Function [2])

Let the most concentrated internal&external stable equilibrium in the cartel formation game, equilibrium binding agreement in the cartel formation game, CPNE in the open membership game, exclusive membership Δ-game and exclusive membership Γ-game, subgame-perfect equilibrium in the sequential move unanimity game and the equilibrium binding agreement in the equilibrium binding agreement game be denoted by M^*(CFG), M^*(C-EBAG), M^*(OMG), M^*(EMΔ-G), M^*(EMΓ-G), M^*(SMUG) and M^*(EBAG) respectively. Then the following relations hold:

Generally:

a) *M^*(CFG)$\pi(\pi)$ M^*(C-EBAG).*

b) *M^*(CFG)πM^*(OMG)πM^*(EMΔ-G)πM^*(EBAG), M^*(CFG)$\pi\pi M^*$(OMG), M^*(CFG)$\pi\pi M^*$(EMΔ-G), M^*(OMG)$\pi\pi M^*$(EBAG) and M^*(EMΔ-G) $\pi\pi M^*$(EBAG) under conditions C_1 to C_3 and the conditions of Propositions 10 and 12 if an OMG exists.*

c) *M^*(EMΓ-G) is weakly more concentrated (coarser) than any equilibrium in any other coalition game under condition C_1.*

For payoff function [2] and N≥5:

$$d)\quad M^{*}(CFG)\pi M^{*}(OMG)\pi M^{*}(EM\Delta\text{-}G)\pi \left\{ \begin{array}{l} M^{*}(C-EBAG) \\ M^{*}(EBAG) \\ M^{*}(SMUG) \end{array} \right\} \pi M^{*}(EM\Gamma\text{-}G),$$

$$e)\quad M^{*}(C\text{-}EBAG)\pi\phi M^{*}(SMUG),\ M^{*}(C\text{-}EBAG)\pi\phi\ M^{*}(EBAG).$$

Proof: Statement a follows by construction of a CFG and C-EBA. All other statements follow immediately from previous propositions and Table 3 (**Q.E.D.**).

Table 3: Equilibrium Coalition Structures for Payoff Function [2]*

N	CFG	C-EBAG	OMG	EMΔ-G	EMΓ-G.	SMUG	EBAG
2	*(2)*	*(2)*	*(2)*	*(2)*	*(2)*	*(2)*	*(2)*
3	*(3)*	*(3)*	*(3)*	*(3)*	*(3)*	*(3)*	*(3)*
4	*(3, 1)*	*(3, 1)*	(2, 2)	*(3, 1)*, (2, 2)	*(4), (3, 1)*	*(3, 1)*	*(3, 1)*, (2, 2)
5	(3, 1, 1)	*(5)*	(3, 2)	(3, 2)	*(5), (4, 1)*	*(5)*	*(5)*
6	(3, 1, 1, 1)	*(5, 1)*	(3, 3)	(3, 3)	*(6), (5, 1)*	*(5, 1)*	*(5, 1)*, (3, 3)
7	(3, 1, 1, 1, 1)	(5, 1, 1)	(3, 2, 2)	(3, 3, 1), (3, 2, 2)	*(7), (6, 1), (5, 2)*	*(5, 2)*	*(5, 2)*
8	(3, 1, 1, 1, 1, 1)	*(8)*	(3, 3, 2)	(3, 3, 2)	*(8), (7, 1), (6, 2)*	*(8)*	*(8)*
9	(3, 1, 1, 1, 1, 1, 1)	*(8, 1)*	(3, 3, 3)	(3, 3, 3)	*(9), (8, 1), (7, 2)*	*(8, 1)*	*(8, 1)*
10	(3, 1, 1, 1, 1, 1, 1, 1)	(8, 1, 1)	(3, 3, 2, 2)	(3, 3, 3, 1), (3, 3, 2, 2)	*(10), (9, 1), (8, 2), (7, 3)*	*(8, 2)*	*(8, 2)*, (5, 5)
11	(3, 1, 1, 1, 1, 1, 1, 1, 1)	(8, 1, 1, 1)	(3, 3, 3, 2)	(3, 3, 3, 2)	*(11), (10, 1), (9, 2), (8,3)*	*(8, 3)*	*(8, 3)*
12	(3, 1, 1, 1, 1, 1, 1, 1, 1, 1)	*(12)*	(3, 3, 3, 3)	(3, 3, 3, 3)	*(12), (11, 1), (10, 2), (9, 3), (8, 3, 1)*	*(8, 3, 1)*	*(8, 3, 1)*, (6, 6)

* Notation as in the text. Under the cartel formation game and the cartel formation game with equilibrium binding agreement only the coarsest equilibrium coalition structures are listed. Under the open membership game, the exclusive membership Δ-game and the exclusive membership Γ-game coalition-proof equilibrium coalitions are listed. The italic coalition structures constitute a Pareto-optimal coalition structure. Under the equilibrium binding agreement only the coarsest equilibrium coalition structures are displayed.

1) CFG versus C-EBAG

In section 3 we started out by deriving the equilibrium in the cartel formation game (CFG). This game exogenously restricts the number of non-trivial coalitions to one. We then introduced the aspect of farsightedness into this game, which implied that a larger coalition could be sustained (Proposition 19a). That is, if countries are aware that if they were to take a free-ride by leaving an IEA others would follow suit and therefore refrain from doing so, larger IEAs can be sustained. Consequently, farsightedness will generally lead to lower global emissions and for payoff function [2] to higher global welfare. Thus from a normative point of view one may hope that negotiators in global pollution control do not take a myopic view. This result is also reminiscent to the dynamic games models discussed in the Introduction.

2) CFG versus OMG and EMΔ-G

It has also been pointed out that the equilibrium in the CFG implies de facto an open-membership rule. Moreover, is has been laid out that allowing for exclusivity in the CFG has no effect on the equilibrium (under the conditions of Proposition 8). Recalling that in the CFG, the open membership game (OMG) as well as in the exclusive membership Δ-game (EMΔ-G) equilibrium coalition structures must be stand-alone stable, the difference of equilibrium coalition structures mainly stems from the exogenous assumption of a single non trivial coalition under the first game and the possibility of multiple coalitions under the latter two games. Thus, the predictions of the CFG have been a conservative and pessimistic estimate. Obviously, allowing for multiple coalitions implies more concentrated equilibrium coalitions, which can also be compared under coarsering. For the relation between M^*(CFG) and M^*(OMG) this is true for payoff function [2] (Proposition 19d) and/or for the assumptions of Proposition 10 (Proposition 19b). However at a more general level, it may well be the case that, say, c=(3, 1, 1, 1) is stand-alone stable but c=(3, 2, 1) and c=(2, 2, 2) are not and therefore under the OMG only c=(2, 2, 1, 1) may be a CPNE. A comparison between the CFG and the EMΔ-G relies on less specific assumptions. Since all stand-alone stable coalition structures are NE in the EMΔ-G, M^*(CFG)π(π) M^*(EMΔ-G) is generally true (Proposition 19b).

3) OMG, EMΔ-G and EMΓ-G

A distinguishing feature among the games with simultaneous choice of membership (allowing for multiple coalitions) is whether membership is open or exclusive and how exclusivity is defined. In the global emission game with symmetric players exclusivity leads to more concentrated coalition structures (Proposition 19b and c). The reason is the following. Generally, countries prefer to be a member of the smallest coalition for a given coalition structure (C_1).

More specifically, countries of larger coalitions may have an incentive to join smaller coalitions if the gain from taking on less climate responsibility is larger than the increase damages due to an increase in global emissions (C_4). Whereas in the OMG countries do not need the consent of members of smaller coalitions, they have to ask for permission in the EMΔ-G. Put differently, the requirement of unanimous agreement within a coalition fosters a more concentrated coalition structure. The fact that the most concentrated coalition structure under the EMΓ-G is more concentrated (coarser) than under the EMΔ-G (Proposition 19c) is due to the more restrictive definition of consent among coalition members in the former game. Once countries deviate in the EMΓ-G the remaining coalitions to which the belonged break apart, though it may be beneficial for them to stick together. Thus deviation is punished hard.

4) EMΔ-G versus EBAG

Whereas in the EMΔ-G the membership rule is spelled out explicitly, an equilibrium binding agreement (EBA) implies de facto also an exclusive membership rule. Recalling the recursive definition of the concept – a coalition c is an EBA if and only if there is no coalition structure that is a refinement and blocks c – forming a coalition in c from the refinement requires unanimous consent of the countries involved in the move. Moreover recall that for C_1 to C_3 any stand-alone stable coalition structure is an EBA and that stand-alone stability is a necessary and sufficient condition for a NE in the EMΔ-G. Consequently, M^*(EMΔ-G) $\pi(\pi)\,M^*$(EBAG) (Proposition 19b) is mainly due to the farsightedness reminiscent to the EBAG. Though conclusions have to be drawn with caution since the EMΔ-G assumes a simultaneous coalition formation process whereas the EBAG a sequential process (and we assumed symmetric countries), the result again (see comparison 1 above) suggests that farsightedness is conducive to the formation of IEAs with respect to global emissions and for payoff function [2] also to global welfare.

5) C-EBAG versus EBAG, C-EBAG versus SMUG and EBAG versus SMUG

From comparison 2 above one would expect that allowing for multiple coalitions in an EBAG implies more concentrated coalitions than if the number of non trivial coalitions is restricted to one (C-EBAG). However, surprisingly, this conjecture is not confirmed (Proposition 19e). Obviously, no clear-cut relation can be established. For payoff function [2] and N=11 M^*(C-EBAG)$\pi\,M^*$(EBAG) and for N=12 this relation is reversed. Thus, also no conclusions with respect to global welfare and global emissions are possible. Also for the equilibrium binding agreement in the cartel formation game (C-EBAG) and the sequential move unanimity game (SMUG) no ranking according to concentration can be established (Proposition (19e). As pointed out in section 3, since we have not yet been able to completely characterize EBAs, a comparison with the SMUG has to be conducted in the future. However from our preliminary results, one would expect that for payoff function [2] $M^*(SMUG) \subset M^*(EBAG)$. Nevertheless, it has been shown that there are EBAs that imply a higher global welfare (and lower global emissions) than the equilibrium in the SMUG. Whether this relation may also be reversed is an open question. For payoff function [2] this seems not to be the case. However, for asymmetric countries this will most likely be true.

6) SMUG versus EMΓ-G

As pointed out in section 3, also the SMUG assumes the strong form of unanimous agreement among the members who form a coalition. Each prospect member who is asked by the initiator whether he likes to join knows the set of all prospective members and can always turn down a proposal. Similar to the EMΓ-G, if and only if all prospective members agree to form a coalition, the coalition will form. In contrast to the EMΓ-G, where dissent implies that coalitions break apart, in the SMUG a prospective member can initiate a new proposal if she does not accept a proposal. Thus, apart from this difference, both games basically differ only with respect to the timing of the formation process. Whereas for the assumption of simultaneous moves no unique equilibrium emerges for payoff function [2], the assumption of sequential moves picks one particular equilibrium out of them. Though among the set of most concentrated coalition structures the relation between both games is clear (Proposition 19c) since in the EMΓ-G the grand coalition is always a CPNE, it is a priori not evident whether the grand coalition will actually come about. In fact, due to C_1, one should expect that it is rather unlikely that the grand coalition will actually materialize. In such cases the equilibrium in the SMUG may be more concentrated than in the EMΓ-G (see Table 3).

7) EMΓ-G versus all other games

The strong assumption of perfect unanimous consent among coalition members implies that the grand coalition is always a CPNE (and SNE) in the EMΓ-G, which by definition is the most concentrated coalition structure.

4.2 Discussion

4.2.1 Single versus Multiple Coalitions

Considering the most concentrated equilibrium coalition structures among the games CFG, OMG, EMΔ-G and EMΓ-G (Proposition 19b and c) with simultaneous choice of membership our results suggest that the prediction of the CFG has been very pessimistic. Among the games with sequential choice of membership such a relation could not be established. Given the record of past IEAs, which basically constitute only one non trivial coalition, the following question comes to mind: Are existing IEAs an equilibrium resulting from a) unrestricted coalition formation or b) have institutional and/or political restrictions been imposed which lead to these coalition structures? If answer a) is correct this would suggest that the results established for simultaneous move games may not hold for asymmetric countries and/or that actual formation processes of IEAs are sequentially. If answer b) is correct this would suggest that it has not been a good idea to focus on a single coalition in the past if coalition formation takes place simultaneously. For instance, in the Kyoto Protocol the US advocates that also developing countries should participate in the agreement. Thus, it may well be the case that more could be achieved in terms of global emission reduction and global welfare if separate agreements would be signed, say among industrialized countries, developing countries and countries in transition. Of course, we do not know which aspects apply in reality. The main uncertainty stems from the fact that in reality countries are heterogeneous. However, so far, assuming heterogeneity has produced no general results. Barrett (1997b) assumed two groups of countries with different characteristics and ran simulations in a CFG. He basically confirmed that the equilibrium coalition will be rather small and will not much improve upon the status quo. Botteon/Carraro (1997 and 1998) ran also simulations in a CFG for a data set of five world regions. They showed that the members of the equilibrium coalition depend on the bargaining rule according to which the coalition members share the gains from cooperation. Moreover, there are different possibilities which members form the equilibrium coalition. Thus in future research it would be interesting either to use Barrett´s and/or Botteon/Carraro´s assumption of heterogeneity or some other assumption and look at the difference between the equilibrium coalition structures in CFG and in other coalition games. In order to get some robust estimates this should be done for different assumptions with respect to the allocation rule of the gains from cooperation. We suspect that also for heterogeneous countries allowing for multiple coalitions may lead to better results in terms of global emissions and welfare. For heterogeneous countries it might be easier to agree on ambitious abatement targets if countries with relatively homogenous interest form coalitions instead of concentrating on only one coalition. Moreover, for heterogeneous countries the overall success will crucially depend on the allocation rule of the gains from cooperation. Two aspects are important. First for a given sharing rule it may not necessarily be optimal if coalition members maximize aggregate welfare of the coalition. Second, if we assume that coalition members maximize aggregate welfare we may search for an optimal allocation rule that generates a coalition structures with highest possible global welfare. Inspirations in this direction may be found in Chander/Tulkens (1997) in the context of the core.

4.2.2 Farsightedness versus Myopic Behavior

Comparisons 1 and 4 in subsection 4.2.1 suggest that if negotiators take a long-term instead of a myopic view to the problem of global pollution this will have a positive effect on global emission reduction. This result is in line with intuition and is also found in the literature on dynamic game models. It therefore seems quite robust. For future research it would be interesting to characterize the conditions under which this conclusion can also be drawn with respect to global welfare. Once more, the main problem will be to establish this result for heterogeneous countries.

4.2.3 Simultaneous versus Sequential Coalition Formation

A comparison between simultaneous and sequential formation games is flawed by three caveats. First, for the sequential choice of membership not very general results can be derived. Second, some of the games are not directly comparable since they differ in more than in the feature of timing. Third, the EMΓ-G applies a very restrictive definition of exclusivity (see subsection 4.3). Given these caveats, and dropping the EMΓ-G from our sample, sequential games lead to more concentrated equilibrium coalition structures than simultaneous games. However, a closer look at the concepts reveals that in the sequential games C-EBAG, EBAG and SMUG farsightedness is an important feature and that this feature is missing in the simultaneous games we investigated. Thus, it seems obvious that differences are not due to the timing of the formation process but due to the aspect of farsightedness and our discussion in subsection 4.2.2 applies. Thus for future research it would be interesting to introduce the aspect of farsightedness in the simultaneous games as for instance as in Chwe´s (1994) farsighted coalitional stability equilibrium.

4.2.4 Exclusive versus Open Membership

For a fair comparison it seems sensible to concentrate only on simultaneous games. As comparisons 3 and 7 in subsection 4.1 clearly indicate at a general level, exclusivity leads to more concentrated coalition structures and is therefore preferable from an ecological perspective. With respect to welfare this can only be established for the EMΓ-G in relation to all other games at a general level. However for payoff function [2], a ranking leads to clear-cut results. Abstracting from the underlying technical forces, which have been laid out extensively in subsection 4.1, we find this result counterintuitive with an interesting political and normative implication. The result is counterintuitive since one should expect that in global pollution control countries which have taken on climate responsibility like other countries to join them. Moreover, unanimous agreement among members, as applies in exclusive membership games, should make it more difficult to fight global pollution than in open membership games. However, it turns out that just the opposite is true, suggesting that the frequently observed and perceived obstacle in international politics, namely to agree by consent, may in fact be an advantage. Moreover, for practical purposes environmental treaties should not generally be open to every candidate. In other words, a public good agreement should be to some extent turned into a club good agreement. In future research it would be interesting to find out whether and under which conditions such a relation can also be established for heterogeneous countries.

4.3 Evaluation

In this subsection we briefly want to review conceptual aspects of some coalition formation games.

1) In the OMG, EMΔ-G and the EBΓ-G the amount of NE is usually quite large. Therefore the concept of CPNE is quite useful in selecting an equilibrium (equilibria). However, the SNE concept may define coalition stability to narrowly. Thus, there is no SNE in the OMG and the EMΔ-G for $N \geq 5$ and payoff function [2], though in the EBΓ-G, SNE and CPNE coincide under the conditions of Proposition 14. As pointed out in section 3, the weakness of the CPNE concept is that it only considers deviations following an initial deviation. In contrast, the EBA allows for the possibility that after a deviation of leading perpetrators has occurred, secondary perpetrators may also deviate if this in their interest. However, this additional and import aspect of the EBAG is paid for by the disadvantage that only members of the same coalition can jointly deviate by forming smaller coalitions. Recall, starting from the grand coalition or some other coalition structure, deviation always implies that coalitions split apart leading to a finer coalition structure.

2) The definition of exclusivity in the EMΓ-G seems very restrictive. It implies that players decide according to having an "ideal" coalition or no coalition at all. That is, once a country or group of countries leaves an existing coalition the remaining coalition(s) to which the deviating countries belonged break apart. This implicit threat is very harsh and appears to be not very credible. If it is beneficial for the remaining countries to stay together why should they split up into singletons?[29]

[29] In the context of dynamic game models such a threat would be called subgame-perfect but not renegotiation-proof. A similar strong assumption is made in the reduced stage models in which the core is applied. See the Introduction.

3) In terms of realistic modeling it seems that depicting the formation of IEAs as a sequential process comes closer to what is actually going on in politics than assuming a simultaneous formation process. Casual evidence suggests that a group of initiators (green countries) kick of the process and look for other countries to join them. As it seems, this process is frequently initiated by only one group of countries, which suggests that the SMUG and the EBAG are a good choice to model the formation of IEAs. However, intuitively, one would expect that the initiators are those countries which form large coalitions whereas in the model just the opposite is true. This putative contradiction is easily resolved by recalling that due to the general condition C_1, countries like to be a member of the smallest coalition. Thus, there is a first mover advantage by announcing to remain a singleton or to be a member of a small coalition, thereby free-riding on the abatement efforts of other countries.

We have a more serious concern with the motivation of the EBAG which we find lacks appeal. We are not aware of any formation process where all countries gathered initially. Also from a conceptual point this assumption exhibits a weakness. Suppose payoff function [2] and N=12 for which we find $\tilde{C}^{EBA}(12) = \{(8,3,1),(6,6)\}$. If the formation actually starts from the grand coalition, then we should expect (8, 3, 1) due to $\pi_i(1,(8,3,1)) > \pi_i(6,(6,6))$. That is, it is better for a perpetrator deviating by herself than to ask other countries to follow suit. Thus, we find it more convincing to view the formation process as described by the algorithm: countries start from the singleton coalition structure and gradually form larger coalition until no further enlargement is self-enforcing.

4) Our results indicate that a grand coalition cannot be expected in the CFG, OMG and EMΔ-G, may occasionally come about in the C-EBAG, EBAG and the SMUG (depending on the number of countries) and is always among the set of equilibrium coalitions in the EMΓ-G. This may be seen as an encouraging result. However, optimism is derived from the questionable restrictive assumption of exclusivity in the EMΓ-G and due to the "heroic" assumption of farsightedness in the C-EBAG, EBAG and SMUG. Taken these qualifications into consideration, we generally have to reckon with suboptimal equilibrium coalition structures in global pollution control. The reason is the free-rider incentive in international pollution control. As pointed out in the Introduction, all formation games discussed in this paper belong to reduced stage game which capture only the first aspect of free-riding, namely that of taking on no or less climate responsibilities than other countries and to benefit from their activities. The second aspect of free-riding where a country joins a coalition but does not fulfill its obligations is not captured. This is evident by considering the following proposition.

Proposition 20: Equilibrium Coalition Structure for N=2 in All Coalition Formation Games

For N=2 the grand coalition is

a) the most concentrated stand-alone stable coalition structure and

b) an equilibrium coalition structure in all analyzed coalition formation games.

Proof: a) follows from $\pi_1(2,(2)) \geq \pi_1(1,(1,1))$ and (2) is more concentrated than (1, 1). b) follows immediately from a) (**Q.E.D.**).

Thus for N=2 the first free-rider incentive is not present since for symmetric countries the grand coalition always dominates the Nash equilibrium. Consequently, in all coalition games the grand coalition is an equilibrium. However, also for N=2 the global emission game resembles a prisoners´ dilemma or least a chicken game (Finus 2001, ch. 9) where conventional theory predicts a suboptimal outcome. Thus, when interpreting the results above one should always keep in mind that if the second aspect of free-riding plays some role, predictions would be more pessimistic.

5. Literature

Aumann, R. (1959), Acceptable Points in General Cooperative N-Person Games. In Contributions to the Theory of Games IV. Princeton University Press, Princeton.

Barrett, S. (1991), The Paradox of International Environmental Agreements. Mimeo, London Business School, London.

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. (1997a), Toward a Theory of International Environmental Cooperation. In: Carraro, C. and D. Siniscalco (eds.), New Directions in the Economic Theory of the Environment. Cambridge University Press, Cambridge, ch. 8, pp. 239-280.

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

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

Bernheim, D., B. Peleg and M. D. Whinston (1987), Coalition-Proof Nash Equilibria. I. Concepts. "Journal of Economic Theory", vol. 42, pp. 1-12.

Black, J., M. D. Levi and D. de Meza (1992), Creating a Good Atmosphere: Minimum Participation for Tackling the "Greenhouse Effect". "Economica", vol. 60, pp. 281-293.

Bloch, F. (1995), Endogenous Structures of Associations in Oligopolies. "RAND Journal of Economics", vol. 26, pp. 537-556.

Bloch, F. (1996), Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division. "Games and Economic Behavior" vol. 14, pp. 90-123.

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.

Botteon, M. and C. Carraro (1998), Strategies for Environmental Negotiations: Issue Linkage with Heterogenous Countries. In: Hanley, N. and H. Folmer (eds.), Game Theory and the Global Environment. Edward Elgar, Cheltenham et al., ch. 9, pp. 180-200.

Carraro C. and F. Moriconi (1997), International Games on Climate Change Control. Working Paper, September 1997, Fondazione Eni Enrico Mattei, Milano.

Carraro, C. and D. Siniscalco (1991), Strategies for the International Protection of the Environment. Working Paper, March 1991, Fondazione Eni Enrico Mattei, Milano.

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 (1991), Strategically Stable Cost Sharing in an Economic-Ecological Negotiation-Process. CORE Discussion Paper No. 9135, Center for Operations, Research and Econometrics, Université Catholique de Louvain, Louvain

Chander, P. and H. Tulkens (1992), Theoretical Foundations of Negotiations and Cost Sharing in Transfrontier Pollution Problems. "European Economic Review", vol. 36, pp. 388-398.

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.

Chatterjee, K. et al. (1993), A Noncooperative Theory of Coalitional Bargaining. "Review of Economic Studies", vol. 60, pp. 463-477.

Chwe, M. S.-Y. (1994), Farsighted Coalitional Stability. "Journal of Economic Theory", vol. 63, pp. 299-325.

D´ Aspermont and J. J. Gabszeweiz (1986), On the Stability of Collusion. In: Matthewson, G. F. and J. E. Stiglitz (eds.), New Developments in the Analysis of Market Structure. Macmillian, New York, pp. 243-264.

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

Farrell, J. and E. Maskin (1989), Renegotiation in Repeated Games. "Games and Economic Behavior", vol. 1, pp. 327-360.

Finus, M. (2000), Game Theory and International Environmental Cooperation: A Survey with an Application to the Kyoto-Protocol. Working Paper No. 86.2000, Fondazione Eni Enrico Mattei, Milan.

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

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

Germain, M., P. L. Toint and H. Tulkens (1996), Calcul économique itératif et stratégique pour les négociations internationales sur les pluies acides entre la Finlande, la Russie et l´Estonie. "Annales d´Économie et Statistique", vol. 43, pp. 101-127.

Germain, M., P. L. Toint, H. Tulkens and A. de Zeeuw (1998), Transfers to Sustain Core-Theoretic Cooperation in International Stock Pollutant Control. CORE Discussion Paper No. 9832, Center for Operations Research and Econometrics, Université Catholique de Louvain, Louvain.

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.

Kalai, E., (1977), Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons. "Econometrica", vol. 45, pp. 1623-1630.

Kalai, E. and M. Smorodinsky (1975), Other Solutions to Nash´s Bargaining Problem. "Econometrica", vol. 43, pp. 513-518.

Nash, J. (1950), Equilibrium Points in N-Person Games. "Proceedings of the National Academy of Sciences", vol. 36, pp. 48-49.

Ray, D. and R. Vohra (1997), Equilibrium Binding Agreements. "Journal of Economic Theory", vol. 73, pp. 30-78.

Ray, D. and R. Vohra (1999), A Theory of Endogenous Coalition Structures. "Games and Economic Behavior", vol. 26, pp. 286-336.

Rubinstein, A. (1982), Perfect Equilibrium in a Bargaining Model. "Econometrica", vol. 50, pp. 97-109.

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. (1979), An Economic Model of International Negotiations Relating to Transfrontier Pollution. In: Krippendorff, K. (ed.), Communication and Control in Society. Gordon and Breach, New York, pp. 199-212.

von Neumann, J. and Morgenstern, O. (1944), The Theory of Games and Economic Behavior. Wiley, New York.

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.

6. Appendix

Appendix 1: Proof of Proposition 2

The fact that payoff functions [2] and [3] satisfy condition C_1 and C_2 has been shown in subsection 2.2. Both functions also satisfy condition C_3 since concentration reduces global emissions (Proposition 1c) from $e^{T(1)}$ to $e^{T(2)}$, $e^{T(1)} > e^{T(2)}$, and for coalition c_i, which lost a member joining coalition c_j, we find i: $\beta'(e_i^{(1)}) = c_i\phi'(e^{T(1)}) > (c_i - 1)\phi'(e^{T(2)}) = \beta'(e_i^{(2)}) \Rightarrow e_i^{(1)} < e_i^{(2)}$. Consequently, benefits increase and damages of former members of coalition c_i decrease.

Condition C_4 and C_5 are satisfied for payoff function [2] because of the following facts.

1) Equilibrium emissions of a member of coalition c_k are given by

[A_1] $$e_k(c_k) = \frac{bd - c_k \cdot c}{b} \; .$$

2) Equilibrium emissions of countries which are not involved in changes of the coalition structure remain the same for payoff function [2].

3) Computing payoffs of country k which left coalition i before and after the accession to coalition j delivers:

[A_2] $$\pi_k(c_i, c) - \pi_k(c_j + 1, c') = \frac{c^2(c_i + c_j - 3)(c_j - c_i + 1)}{2b} > 0 \text{ if } 2 \le c_i \le c_j.$$

Similarly, if coalition c_i and c_j merge we find:

[A_3] $$\pi_i(c_i, c) - \pi_i(c_i + c_j, c') = \frac{c^2 \cdot c_j \cdot (c_j - 2 \cdot c_i)}{2b} >\le 0$$

[A_4] $$\pi_j(c_j, c) - \pi_j(c_i + c_j, c') = \frac{c^2 \cdot c_i \cdot (c_i - 2 \cdot c_j)}{2b} < 0 \; .$$

[A_2], [A_3] and [A_4] give rise to conditions C_4 and C_5 respectively.

In order to show that condition C_4 holds for payoff function [3], we proceed in a similar fashion.

1) If a member of a coalition c_j joins a coalition a c_i equilibrium emissions are given by

[A_5] $$e^{T(1)} = \frac{Nd}{1 + c/b(c_i^2 + c_j^2 + CR)}, \; e_j^{(1)}(c_j) = \frac{d(b + (c_i^2 + c_j^2 + CR) \cdot c - c_j cN)}{b + (c_i^2 + c_j^2 + CR) \cdot c}$$

and

[A6] $$e^{T(2)} = \frac{Nd}{1+{}^{c}\!/_{b}((c_i+1)^2+(c_j-1)^2+CR)}\ ,$$

$$e_i^{(2)}(c_i+1) = \frac{d(b+((c_i+1)^2+(c_j-1)^2+CR)\cdot c - c(c_i+1)N)}{b+((c_i+1)^2+(c_j-1)^2+CR)\cdot c}$$

where the superscript 1 indicates the initial situation and 2 the situation when a member of coalition c_j joins coalition c_i, T stands for total emissions, and $CR = \sum_{k\neq i,j} c_k$. Inserting those emissions in payoff function [3] and computing $\pi_j^{(1)} - \pi_i^{(2)}$ delivers a rather big term which, however, can be shown to be negative using $2 \le c_i \le c_j$ (**Q.E.D.**).

Appendix 2: Proof of Proposition 3

A concentration may either involve a) a merger of coalition c_i and c_j or b) a country k of coalition c_i leaves coalition c_i and joins coalition c_j, $c_i \le c_j$. Any concentration follows from a sequence of these actions. Since non-active players benefit from a concentration, it suffices to look at the effect on the "active" players if this is unambigiously positive.

In case a) we find for payoff function [2]:

[A7] $$c_i \cdot (\pi_i(c_i+c_j, c') - \pi_i(c_i, c)) + c_j \cdot (\pi_j(c_i+c_j, c') - \pi_j(c_j, c)) = \frac{c_i c_j c^2 (c_i + c_j)}{2b} > 0$$

and in case b):

[A8] $$\pi_k(c_j+1, c') - \pi_k(c_i, c) + c_j \cdot (\pi_j(c_j+1, c') - \pi_j(c_j, c)) + (c_i - 1)\cdot(\pi_i(c_i-1, c') - \pi_i(c_i, c))$$

$$= \frac{c^2(c_i+c_j)(c_j-c_i+1)}{2b} > 0$$

which obviously can unambigiously be signed (**Q.E.D.**).

Appendix 3: Proof of Proposition 5

For payoff function [2] we compute

[A9] $$\pi_i(c_i, c) - \pi_i(\{i\}, c') = -\frac{c^2(c_i-1)(c_i-3)}{2b}$$

which is positive for $c_i = 2$, zero for $c_i = 3$ and negative for $c_i > 3$.

For payoff function [3] equilibrium emissions are given by

[A10] $$e^{T(1)} = \frac{Nd}{1+{}^{c}\!/_{b}(c_i^2+CR)}\ ,\ e_i^{(1)}(c_i) = \frac{d(b+(c_i^2+CR)\cdot c - c_i c N)}{b+(c_i^2+CR)\cdot c}$$

and

[A11] $e^{T(2)} = \frac{Nd}{1+\frac{c}{b}((c_i-1)^2+CR+1)}$, $e_i^{(2)}(1) = \frac{d(b+((c_i-1)^2+CR+1)\cdot c - cN)}{b+((c_i-1)^2+CR+1)\cdot c}$

where the superscript 1 indicates the initial situation and 2 the situation when a member of coalition c_i leaves the coalition to become a singleton, T stands for total emissions and $CR = \sum_{j\neq i} c_j^{\ 2}$. Inserting those emissions in payoff function [3] delivers

[A12] $\pi_i^{(1)} - \pi_i^{(2)} = -\frac{bd^2c^2N^2(c_i-1)\cdot A}{2(b+cc_i^{\ 2}+cCR)^2(b+cc_i^{\ 2}-2cc_i+2c+cCR)^2}$

$$A := c^2c_i^{\ 5} - 3c^2c_i^{\ 4} + 2c^2c_i^{\ 3}CR + 4c^2c_i^{\ 3} + 2cbc_i^{\ 3} - 6cbc_i^{\ 2} - 2c^2c_i^{\ 2}CR - 4c^2c_i^{\ 2} + b^2c_i$$
$$+2cbc_iCR + c^2c_iCR^2 + 4cbc_i - 4cb - 2cbCR - 3b^2 + c^2CR^2$$

Obviously, if A is positive, then $\pi_i^{(1)} - \pi_i^{(2)} < 0$. A is positive for $c_i > 2$ since: $c^2c_i^5 3c^2c_i^4 \geq 0$, $2c^2c_i^{\ 3}CR - 2c^2c_i^{\ 2}CR > 0$, $4c^2c_i^{\ 3} - 4c^2c_i^{\ 2} > 0$, $b^2c_i - 3b^2 \geq 0$, $2cbc_i^3 - 6cbc_i^2 \geq 0$, $2cbc_iCR$ $-2cbCR > 0$, $4cbc_i - 4bc > 0$. Remark: Since those terms which contain CR are positive for $c_i \geq 1$, the incentive to leave a (non-trivial) coalition and to become a singleton increases in CR which implies a concentration of $c \setminus c_i$ (**Q.E.D.**).

Appendix 4: Proof of Proposition 9c

In order to prove part c of Proposition 9 it suffices to show that the proposition holds for N=$c_{S(i)}$-1 since for this N there are the maximum number of singleton coalitions. If N=$c_{S(i)}$-1 (and recalling Proposition 9b) we have ($c_{S(i)}$-1)-$c_{S(i-1)}$<[1+$\sqrt{2(c_{S(i-1)})^2 - 2c_{S(i-1)}}$]-$c_{S(i-1)}$ singleton coalitions. In the following we show that [1+$\sqrt{2(c_{S(i-1)})^2 - 2c_{S(i-1)}}$]-$c_{S(i-1)}$ $\leq I((c_{S(i)}-1)/3)-1$. Assume the opposite, namly that [1+$\sqrt{2(c_{S(i-1)})^2 - 2c_{S(i-1)}}$]-$c_{S(i-1)}$ $> I((c_{S(i)}-1)/3)-1$ holds. Then

[1+$\sqrt{2(c_{S(i-1)})^2 - 2c_{S(i-1)}}$]-$c_{S(i-1)} > (c_i^S - 1)/3 - 1$

$\Rightarrow$ [1+$\sqrt{2(c_{S(i-1)})^2 - 2c_{S(i-1)}}$]-$c_{S(i-1)}$>([1+$\sqrt{2(c_{S(i-1)})^2 - 2c_{S(i-1)}}$]$-1)/3-1$

$\Rightarrow$ 3[1+$\sqrt{2(c_{S(i-1)})^2 - 2c_{S(i-1)}}$]-3 $c_{S(i-1)}$>[1+$\sqrt{2(c_{S(i-1)})^2 - 2c_{S(i-1)}}$]-1-3

$\Rightarrow$ 2[1+$\sqrt{2(c_{S(i-1)})^2 - 2c_{S(i-1)}}$]>3 $c_{S(i-1)}$-4 $\Rightarrow$ 2 $\sqrt{2(c_{S(i-1)})^2 - 2c_{S(i-1)}}$ >3 $c_{S(i-1)}$-6

$\Rightarrow$ 4 $(2(c_{S(i-1)})^2 - 2c_{S(i-1)})$>9($c_{S(i-1)}$)^2-36 $c_{S(i-1)}$+36

$\Rightarrow$ 4 $(2(c_{S(i-1)})^2 - 2c_{S(i-1)})$>9($c_{S(i-1)}$)^2-36 $c_{S(i-1)}$+36

$\Rightarrow$ ($c_{S(i-1)}$)^2-28 $c_{S(i-1)}$+36<0 $\Rightarrow$ ($c_{S(i-1)}$-14)^2–160<0 $\Rightarrow$ $c_{S(i-1)}$< $\sqrt{160}$+14 ≈ 26.65.

Hence the proposition holds for $c^*_{S(i)} \geq 55$. For $c^*_{S(i)} \leq 38$ the proof follows from evaluation of ($c_{S(i)}$-1)-$c_{S(i-1)}$, e.g., (38-1)-26=11≤I(37/3)-1=12.

Appendix 5: Proof of Proposition 17

As a preliminary information to establish part a)-c) in Proposition 17, we establish the following three lemmas.

Lemma 1A

For the Pareto dominance numbers as defined in Proposition 17, a) $f_i \leq f_{i-1} + f_{i-2}$ and b) $f_i \leq 2 f_{i-1}$ hold.

Proof: Follows by induction. a) We have $f_0 = 1$, $f_1 = 2$, $f_2 = 3 \leq 1 + 2$ and therefore presume $f_j \leq f_{j-1} + f_{j-2} \;\; \forall \;\; j < i$. Thus, we have to show that $f_i \leq f_{i-1} + f_{i-2}$. Assume the opposite, namely, that $f_i > f_{i-1} + f_{i-2}$ would be true. Then $\Phi(f_{i-1} + f_{i-2}) = \{f_{i-1}, \Phi(f_{i-2})\} = \{f_{i-1}, f_{i-2}\}$ (since $f_{i-1} + f_{i-2}$ is not a Pareto dominance number by assumption and f_{i-1} is the number following f_i, $f_i > f_{i-1}$). Consequently, $\{f_{i-1}, f_{i-2}\}$ is not weakly Pareto-dominated by the grand coalition which, using C_5, may be written as $2f_{i-2} < f_{i-1}$ from which $f_{i-1} > f_{i-2} + f_{i-3}$ follows since $f_{i-2} > f_{i-3}$. However, $f_{i-1} > f_{i-2} + f_{i-3}$ contradicts the initial assumption of induction $f_j \leq f_{j-1} + f_{j-2} \;\; \forall \;\; j < i$. From $f_i \leq f_{i-1} + f_{i-2}$ and $f_{i-2} < f_{i-1}$ b) $f_i \leq 2f_{i-1}$ follows (**Q.E.D.**).

Lemma 2A

For the Pareto dominance numbers as defined in Proposition 17, $f_i \geq 3/2\, f_{i-1}$ holds.

Proof: From Lemma 1A we have $f_i \leq f_{i-1} + f_{i-2}$. Two cases may be distinguished.

Case 1: Suppose $f_i = f_{i-1} + f_{i-2}$. Then $f_i \geq \frac{3}{2} f_{i-1}$ is equivalent to $f_{i-1} + f_{i-2} \geq f_{i-1} + \frac{1}{2} f_{i-1}$ and hence equivalent to $2f_{i-2} \geq f_{i-1}$. The last inequality is true because of Lemma 1A, part b.

Case 2: Suppose $f_i < f_{i-1} + f_{i-2}$. Then $\tilde{\Phi}(f_i) := \{f_{i-1}, \Phi(f_i - f_{i-1})\}$ is weakly Pareto-dominated by f_i which may be written as $f_i - f_{i-1} \geq \frac{1}{2} f_{i-1}$ due to C_5. However, $f_i - f_{i-1} \geq \frac{1}{2} f_{i-1}$ $\Leftrightarrow f_i \geq \frac{3}{2} f_{i-1}$ (**Q.E.D.**).

Lemma 3A

$f_i \in \Phi(N) \Rightarrow f_{i-1} \notin \Phi(N)$.

Proof: Without restriction of generality we only have to show that if f_i is the largest PD-number of $\Phi(N)$, then $f_{i-1} \notin \Phi(N)$. Suppose $f_{i-1} \in \Phi(N)$ would be true. Then

$f_{i+1} \le f_i + f_{i-1} \le f_i + (N - f_i) = N$ follows from Lemma 1A which implies the obviously wrong conclusion $f_{i+1} \in \Phi(N)$ (**Q.E.D.**).

We now prove part a) of Proposition 17, namely that the equilibrium coalition structure in the SMUG is equal to the Pareto dominance decomposition for payoff function [2]. We proceed by induction.

Start of Induction

For N=1, we have $\Phi(1) = \{1\} = c^*$.

Assumption of Induction

For all $m < N$, $c^*(m) = \Phi(m)$. Due to condition C_1, the first initiator will propose f^M (and the second f^{M-1} and so on), where $\Phi(m) = \{f^1, ..., f^M\}$ and $f^k > f^{k+1}$ $\forall$ k<M.

Conclusion of Induction

Demonstrate that $c^*(N) = \Phi(N)$ where $\Phi(N) = \{f^1, ..., f^M\}$ with $f^k > f^{k+1}$ $\forall$ k<M is the PD-decomposition derived from the algorithm described in Proposition 17.

Remark

In order to demonstrate this, we have to show that country 1 proposes f^M. Then for the remaining countries the induction assumption can be applied. That is, these countries form coalitions according to the decomposition $\Phi(N - f^M) = \Phi(N) \setminus \{f^M\}$. That is, if f^M players have left the game, the game is the same as it would be played among $N - f^M$ players. This is true since payoff function [2] implies orthogonal reaction functions and thus the emission level of the coalition formed by f^M players is irrelevant for the decision of the remaining players.

Proof

To show that country 1 proposes f^M, suppose the opposite, namely, a proposal $\tilde{c} \ne f^M$.

$\tilde{c} > f^M$:

Case 1: $\tilde{c} < N$, $\Phi(\tilde{c}) \ne \{\tilde{c}\}$

If it were an equilibrium strategy to propose $\tilde{c}$, it would be accepted and the equilibrium coalition structure would be given by $\{\Phi(N - \tilde{c}), \tilde{c}\}$. However, according to the assumption of induction, this cannot be an equilibrium since the coalition $\tilde{c}$ will break up due to $\Phi(\tilde{c}) \ne \{\tilde{c}\}$. (Example: N=14: $c^* = \{8, 5, 1\}$. Suppose $\tilde{c} = 4$, then $\Phi(\tilde{c}) = \{3, 1\}$).

Case 2: $\tilde{c} < N$, $\Phi(\tilde{c}) = \{\tilde{c}\}$, $\tilde{c} \in \Phi(N)$

Due to $\Phi(\tilde{c}) = \{\tilde{c}\}$, the resulting coalition structure would be $\{\Phi(N-\tilde{c}), \tilde{c}\}$. However, since $f^M \in \Phi(N-\tilde{c}) = \Phi(N) \setminus \{\tilde{c}\}$ (or alternatively, $\tilde{c} = f^k > f^M$ where $f^k \in \Phi(N)$), $\pi_1(\tilde{c}, \{\Phi(N-\tilde{c}), \tilde{c}\}) < \pi_1(f^M, \Phi(N))$ follows from C_1. That is, it does not pay country 1 to propose a larger coalition if the final coalition structure will contain a smaller coalition. (Example: N=14: $c^* = (8, 5, 1)$. Suppose $\tilde{c} = 5$, then $\Phi(\tilde{c}) = (8, 5, 1)$ but $\pi_1(5, (8, 5, 1)) < \pi_1(1, (8, 5, 1))$.

Case 3: $\tilde{c} < N$, $\Phi(\tilde{c}) = \{\tilde{c}\}$, $\tilde{c} \notin \Phi(N)$

Again, due to $\Phi(\tilde{c}) = \{\tilde{c}\}$, the resulting coalition structure would be $\{\Phi(N-\tilde{c}), \tilde{c}\}$. Country 1 is worse off by proposing $\tilde{c}$ instead of f^M since 1) $\tilde{c} > f^M$ and 2) $\Phi(N)$ is more concentrated than $\{\Phi(N-\tilde{c}), \tilde{c}\}$. That is, country 1 would 1) *not* be a member of the smallest coalition and, additionally, 2) the resulting coalition structure would be less concentrated. The welfare implications follow from C_1 and C_2. The fact that $\{\Phi(N-\tilde{c}), \tilde{c}\}$ is less concentrated than $\Phi(N)$ is simply an implication of the definition of $\Phi(N)$ (stating that $\Phi(N)$ is the most concentrated decomposition comprising PD-numbers which is not weakly Pareto-dominated by the grand coalition). (Example: N=14: $c^* = (8, 5, 1)$. Suppose $\tilde{c} = 3$, then $c = (8, 3, 3)$ but $\pi_1(1, (8, 5, 1)) > \pi_1(3, (8, 3, 3))$.)

Case 4: $\tilde{c} = N$

In order for $\tilde{c} > f^M$ to be possible, the PD-decomposition must comprise at least two elements. However, by the definition of the PD-decomposition such a decomposition is not Pareto-dominated by the grand coalition (otherwise the grand coalition were to form and $\tilde{c} > f^M$ could not be constructed) and hence $\pi_1(N, (N)) < \pi_1(f^M, \Phi(N))$ must hold in this case.

$\tilde{c} < f^M$:

First note that $\Phi(\tilde{c}) = \{\tilde{c}\}$, otherwise the proposal $\tilde{c}$ cannot be an equilibrium as demonstrated above. If $\Phi(\tilde{c}) = \{\tilde{c}\}$, then $c = \{\Phi(N) \setminus \{f^M\}, \Phi(f^M - \tilde{c}), \tilde{c}\}$. Second, let $f^M = f_i$, then $\tilde{c} = f_k$ where $f_k < f_i$ ($f_k = f_{i-1}$ or $f_k = f_{i-2}$ etc.). However, $\tilde{c} = f_{i-1}$ is not possible, since country 1 would be worse off than in coalition f^M. This follows from $f_i \le 2f_{i-1}$ (Lemma 1A) and applying C_5. Third, the decomposition of f^M into $\{\Phi(f^M - \tilde{c}), \tilde{c}\}$ cannot contain two PD-numbers following each other immediately (e.g., f_{i-2}, f_{i-3}) by $f_i \le 2f_{i-1}$ and C_5. Fourth, $f_i \ge \frac{3}{2} f_{i-1}$ which follows from Lemma 2A. Finally, $f_i \ge \frac{3}{2} f_{i-1}$ implies $f_i \ge \frac{9}{4} f_{i-2} > 2f_{i-2}$. That is, if a PD-decomposition exists without two consecutive numbers, this decomposition must be unique and is given by $\tilde{\Phi}(f^M) := \{f^1(f^M - 1), \Phi(f^M - f^1(f^M - 1))\}$ which is, however, Pareto-dominated by f^M by the construction of the PD-numbers (see Proposition 17).

Formally, let $\{\Phi(f^M - \tilde{c}), \tilde{c}\} = \{g^1, ..., g^L\}$ with PD-numbers $g^i > g^{i+1}$, $g^i = f_j$, then $g^{i+1} \leq f_{i+2}$. Consequently, $\pi_1(\tilde{c}, \{\Phi(f^M - \tilde{c}), \tilde{c}\}) \leq \pi_1(g_L, (g^1, ..., g^L)) < \pi_1(f^M, (f^M))$ must hold.

Finally, statement b) and c) follow directly from Lemma 2A and Lemma 3A (**Q.E.D.**).