

**Game Theoretic Research on the Design of  
International Environmental Agreements:  
Insights, Critical Remarks and Future Challenges\***

von

Michael FINUS

Diskussionsbeitrag Nr. 414  
June 2007

Diskussionsbeitrag der Fakultät für  
Wirtschaftswissenschaft der FernUniversität in Hagen

Herausgegeben vom Dekan der Fakultät

\* Parts of this paper have been written while I was a visiting fellow at the Universidade do Algarve, Faro, Portugal. I particularly like to thank Prof. Pedro Pintassilgo for the invitation and the many interesting and inspiring discussions. I have also benefited from numerous discussions with Prof. Alfred Endres from the University of Hagen that have shaped my view of the subject.

## **Abstract**

In recent years, the number of publications that analyze the formation and stability of international environmental agreements (IEAs) using the method of game theory has sharply increased. However, there have also been voices that criticize game theoretic analyses for being based on too abstract models and assuming too much about the rationality of agents. This paper reports on some recent results that shall demonstrate the usefulness but also the limitations of game theory for the analysis of IEAs. It focuses on the relation between different designs and the success of IEAs. Results are illustrated for the climate change problem with the empirical Stability of Coalitions (STACO) Model developed by Dellink et al. (2004). Subsequently, some features of actual treaty-making not considered with this model are discussed with reference to the literature and their importance for future research is highlighted.

**Keywords:** game theoretic research, international environmental agreements, recent results, critical review, agenda for future research

## 1. Introduction

Since the early papers by Barrett (1994), Carraro/Siniscalco (1993), Chander/Tulkens (1992), Hoel (1992) there is a sharply increasing number of publications that analyze the formation and stability of international environmental agreements (IEAs) using game theory.<sup>1</sup> This is not surprising for at least two reasons.

Firstly, game theory is a mathematical method that studies the interaction between agents based on behavioral assumptions about the preference of agents and makes prediction about the outcome of these interactions by applying various equilibrium concepts. Thus, game theory seems to be an ideal tool to study IEAs as they provide a public good with transboundary externalities from which nobody can be excluded. In particular the use of non-cooperative game theory seems suggestive, presuming the absence of a “third party” able to enforce contracts: though global rationality would call for cooperation, individual rationality often stipulates free-riding behavior since there is no supranational institution that can enforce participation in IEA and compliance with agreed obligations.

Secondly, environmental problems with an international dimension become more and more threatening and receive an increasing coverage in politics, the media and the public. Topics range from the loss of biological diversity, the depletion of the ozone layer and more recently the concern about the impacts of global warming. All these environmental problems make joint and coordinate action at a global scale necessary in order to prevent the most severe damages to nature and mankind. However, despite good intentions and many public statements by politicians, progress is often slow: free-riding is still the most important obstacles for successful IEAs.<sup>2</sup> In order to mitigate free-riding, it is important to understand the strategic considerations of the actors causing transboundary environmental externalities for which game theory is an ideal method.

However, there have also been voices that criticize game theoretic analyses for being based on too abstract models and assuming too much about the rationality of agents. As to the first

---

<sup>1</sup> For a survey see for instance Barrett (2003), Finus (2001, 2003a), Folmer/van Mouche (2000), Folmer/de Zeeuw (1999, 2000) and Wagner (2001).

<sup>2</sup> Climate change is a good point in case. The Framework Convention on Climate Change called for actions to prevent severe damages to the climate. However, this treaty is only a declaration of good intentions without specific abatement obligations, which may explain that it has been signed by almost all countries in 1992. In 1997, only 38 states accepted emission ceilings within the Kyoto Protocol. It went not into force until 2002, after the USA had withdrawn their signature and several concessions to the original abatement targets had been granted to various participants. Current emission levels suggest that many signatories will not be able to meet even these modest obligations in the commitment period 2008-12. Other examples which prove that participation and compliance is a problem when dealing with global environmental problems are provided in Barrett (2003) and Finus (2003a), section 2.

point, I would argue that abstraction is a general phenomenon of models in social sciences. On the one hand, abstraction is necessary to unveil the gist of the problem. Moreover, including too many aspects and layers of a problem renders a model too complex for deriving sensible conclusions. On the other hand, it is certainly true that the simplifying assumptions have to be critically checked whether they really capture the heart of the problem under investigation. In the course of such a review, it seems wise to replace simplifying by more realistic assumptions only sequentially in order to be able to handle such complications.

As to the second point, the definition of rationality becomes important. If irrationality means that agents do not base their actions exclusively on economic variables in a narrow sense, I would certainly agree with this critique. After all, models that define rationality in such a narrow sense frequently fail to explain important real world phenomena. However, if irrationality means that agents' behavior is not related to their preferences, which may include a preference for some notion of equity and fairness, I believe a scientific analysis is not possible anymore. In particular, the question arises which assumption should replace rationality and on which grounds it can be founded.

In this paper I report on some selected results that have been obtained with recent developments in game theory. The aim is threefold. Firstly, I would like to demonstrate how models can account for more realistic assumptions and what the insights from such an analysis are. Secondly, I want to show that there is still sufficient room for developing current models further within the boundaries of rationality in a wider sense. Thirdly, I want to identify the limitations of current models, explaining the challenges for future research, hoping to initiate new ideas for improvement.

In the following, I proceed as follows. In section 2, I will first give a brief overview of game theoretic models on IEAs that deal with the problems of participation (membership models) and compliance (compliance models). Then I describe membership models in more detail as this is the class of games to which I will restrict myself in this paper. In section 3, I analyze how the design of IEAs is related to their success using results derived from the *Stability of Coalition Model* (STACO-model). STACO comprises an empirical climate module with twelve world regions and a game theoretic module computing stable coalitions – the technical term for stable environmental agreements. In section 4, I discuss some aspects not considered in STACO but which may prove important for future research.

## **2. The Theory of Coalition Models**

### **2.1 Introduction**

Up to now, there is no coalition model that captures all aspects of coalition formation. This concerns not only some institutional details of treaties and polit-economic aspects of negotiations, but also the fundamentals of the game theoretic analyses of IEAs – the two types of free-rider incentives (Finus 2003a). The first type refers to the incentive not to participate in an IEA; the second type to the incentive not to comply with the obligations agreed upon in an IEA. Membership models focus on the first type of free-rider incentive. Implicitly, they assume that once a country joins an agreement it will comply with the agreed treaty obligations. Thus, compliance is exogenous. In contrast, compliance models concentrate on the second type of free-rider incentive, starting from the exogenous assumption that some coalition has formed and test whether treaty obligations can be enforced with credible threats to sanction non-compliance. Hence, I conclude that the current models are limited in their scope and that there is a need to develop models that capture both types of free-rider incentives.<sup>3</sup>

Moreover, membership and compliance models have in common that they make exogenous assumptions about the institutional details of the design of an agreement. These assumptions concern for instance the choice of the aggregate level of abatement implemented within a coalition, the allocation among its members, the amount and direction of transfer payments among coalition members. Thus, despite many models compare the outcome of different institutional arrangements, one has to be aware that the choice is not endogenized. This is not a critique, as the complexity of these models requires some simplification, but simply an acknowledgment of the limitation of these models. It is also important to point out that almost all models assume the players in the game to be countries. That is, the political decision process within a country is neglected and the representatives of a country de facto behave as benevolent dictators in the sense of welfare economics. I will later comment on this assumption in section 4.

### **2.2 Membership Models**

Two strands of membership models have been distinguished in the literature (Finus 2003a and Tulkens 1998): cooperative and non-cooperative. Contributions using cooperative game

---

<sup>3</sup> At a theoretic level it could certainly be argued that non-compliance of a participant is de facto the same as leaving an agreement in order to become a non-participant. This is demonstrated in Barrett (1999) who compares the conclusion derived from a membership and a compliance model for the same objective functions. However, at an applied level, things are more complicated and non-compliance and non-participation are two different things. De Zeeuw (2005) may be seen as a preliminary attempt to include participation and compliance in one model.

theory are associated with the stability concept of the core; those using non-cooperative game theory are associated with the concept of internal and external stability and modifications of it.<sup>4</sup> The textbook definition suggests that cooperative in contrast to non-cooperative game theory assumes a third party that can enforce an agreement. Following this definition implied that cooperative game theory models would not be suited to model the problems of cooperation in international pollution control. However, I believe this categorization of models is misleading and the implied conclusion is premature for two reasons. Firstly, also the cooperative models test whether coalition members have an incentive to leave the coalition, considering reactions of the remaining players that imply an implicit punishment. For instance, the  $\gamma$ -core assumes that once a group of countries leaves a coalition, this coalition breaks apart and the remaining countries play their non-cooperative abatement strategy. Secondly, also most non-cooperative models make heroic assumptions about the sense for cooperation among coalition members when choosing their abatement strategies (see more on this below.) Moreover, both strands assume full compliance with abatement obligations once a coalition has formed which is typically associated with the notion of cooperative game theory.

Regardless whether cooperative and non-cooperative game theory is an adequate terminology, and ignoring the challenge to find a better one, I think there is a clear indication that the non-cooperative approach is superior for three reasons. Firstly, taking the assumption of a missing supranational authority seriously means to undertake the effort to include as many non-cooperative elements as possible into a game theoretic model on IEAs. In particular, consistency requires that the behavior of countries or their representatives is derived from individual optimization. However, it is important to point out that this does not preclude that countries are interested in cooperation. Even if agents behave selfish in a strict sense, in the presence of externalities, individual optimization requires some form of cooperation or coordination. Moreover, ethical and moral motives can be considered by making it part of individuals' objective functions, which are frequently called payoff functions in the game theoretic literature. Secondly, recent developments in non-cooperative membership models have shown that this approach is more general by separating the rules of coalition formation from stability concepts (Bloch 2003 and Finus 2003a,b). This more general approach also allows replicating the assumptions and results of the cooperative approach (Finus/ Rundshagen 2006a). Thirdly, the potential for explaining real world phenomena of IEAs is much higher for the non-cooperative than for the cooperative approach (as used up to now). By assumption, only the grand

---

<sup>4</sup> The cooperative approach was initiated by Chander/Tulkens (1992) and the non-cooperative by Barrett (1994), Carraro/Siniscalco (1993) and Hoel (1992). Subsequent papers are reviewed in Finus (2003a).

coalition<sup>5</sup> implementing the socially optimal abatement target can be stable in the sense of the core. As most papers (e.g. Chander/Tulkens 1995, 1997) show, this is indeed the case provided an adequate transfer scheme is implemented. The reason is that the implicit punishment implied by the core concept is very strong. Ignoring the discussion whether this punishment is credible or whether alternative assumptions are more convincing, it is evident that this concept cannot much contribute to explain the existing problems of global environmental cooperation. Of course, one may point out that the core is a normative and not a positive concept. However, then the question arises whether we still and repeatedly have to be reminded that the benchmark to solve transboundary externalities is full cooperation. I suppose this is known also to non-game theorists and is not even debated among environmental and ecological economists as well as scientists of other disciplines. What seems controversial is the identification of the problems that prevent full cooperation or even partial cooperation and the measures to overcome these obstacles. A small contribution from a (non-cooperative) game theorists point view is offered below.

### 2.3 The Construction of Non-cooperative Membership Models

Table 1 shows the basic structure of non-cooperative membership models. In the first stage, countries decide upon their participation; in the second stage, they decide upon abatement and possible transfers. The decision about participation leads to a coalition structure which is a partition of countries in disjoint sets.<sup>6</sup> For given assumptions about the decisions in the second stage, roughly speaking, a coalition structure is called stable if no country has an incentive to revise its participation decision.<sup>7,8</sup>

Within each stage, different assumptions can be made. I first describe the archetype of non-cooperative membership models, which I call henceforth the base model, and of which the assumptions are underlined in Table 1. The base model assumes that countries simultaneously

---

<sup>5</sup> This is a coalition of all countries.

<sup>6</sup> This means that a coalition structure comprises of coalitions. A coalition with only a single country is called singleton. Henceforth, I only use the term “coalition” for a non-trivial coalition of at least two countries. In non-technical terms, this is referred to as an agreement. If all countries are singletons, I call this the singleton coalition structure and if all countries form one coalition, I call this the grand coalition.

<sup>7</sup> This corresponds to a Nash equilibrium in participation strategies. In the context of the simultaneous move single coalition open membership game, the archetype of coalition models described below, this stability concept has also been called internal and external stability following d’Aspremont et al. (1983). That is, no member has an incentive to leave the coalition (i.e. agreement) to become a singleton, and no outsider has an incentive to join the coalition. There are also more sophisticated equilibrium concepts. However, I ignore these technical refinements in the following because they are of minor applied relevance. See for instance Bloch (2003) and Finus (2001).

<sup>8</sup> Note that the singleton coalition structure is stable in the sense of a Nash equilibrium by definition. Suppose all countries announce not to join the agreement, then a deviation by a single country (announcing to join the agreement) makes no difference – this requires at least a deviation by two countries.

decide whether to join an IEA. This is a one-shot decision. Hence, even if a model considers a dynamic payoff structure (as for instance the climate models reported in section 3.1), the decision will be based on discounted payoffs. Moreover, the base model assumes that there is only a single agreement (see footnote 7) where those countries that do not join the agreement remain as singletons (frequently called non-signatories in the literature); all other countries are members of the agreement (frequently called signatories in the literature). Nobody can be denied the access to the agreement which is termed open membership.

Once a coalition has formed, it is assumed that all countries simultaneously choose their abatement levels which is termed the Cournot-assumption in the literature. Coalition members choose their abatement as to maximize the aggregate payoff to their coalition, every non-member maximizes its own payoff. This means that coalition members internalize all externalities within their coalition, i.e. abatement is efficient within the coalition, but they ignore the externalities imposed on countries outside their coalition. Specifically, the coalitional abatement level is optimal for the coalition and the allocation among its members is cost-effective.<sup>9</sup> Hence for this assumption if the grand coalition were to form, it would implement socially optimal abatement levels. If no coalition formed (i.e. the singleton coalition structure formed), this would correspond to non-cooperative abatement levels, also frequently called Nash equilibrium. Any coalition comprising more than one but less than all members may be regarded as partial cooperation. The base model assumes no transfers to correct for a possible asymmetric welfare distribution. Moreover, the decisions in the second stage are taken by representatives acting like social planners of individual countries; they are only based on costs and benefits from abatement (material payoffs), these are known with certainty and have been objectively estimated based on current scientific evidence.

In the following, I consider modifications of the base assumptions. In section 3, I concentrate on those alternative assumptions for which results have been obtained with STACO. In section 4, I discuss all remaining alternative assumptions.

### **3. Results from the STACO Model**

#### **3.1 Introduction**

In recent years, a couple of climate models have been used in order to study the prospects of cooperation. Examples include for instance Babiker (2001), Böhringer/Löschel (2006), Buchner/Carraro (2005a), Guo et al. (2006), Kempfert (2005) and Tol (2001). Most of these model are computable general equilibrium (CGE) integrated assessment models that capture

---

<sup>9</sup> With coalitional abatement I mean the aggregate abatement level of the coalition.

the feedback between the economy and damages from climate change. In order to study stability of agreements, these models have to be linked to a game theoretic module of coalition formation. For instance, the Climate World Simulation Model (CWSM) is a modification of the RICE model by Nordhaus/Yang (1996) and stable coalitions are determined with algorithms programmed with the software package GAMS (e.g. Eyckmans/Finus 2006a and Eyckmans/Tulkens 2003).

For obvious reasons, CGE-models have been subjected to many criticism, starting from their assumptions about parameter values, the simply way in which they capture the climate and many other things (Scricciu 2006). I will not challenge this criticism at this stage; in fact will add some other points in section 4. Instead, I want to point out that as long as we are not capable of solving most of the interesting coalition models analytically, we have to rely on simulations. Hence, the only question remains whether simulations should be based on some scientifically estimated parameters according to the current state of the art or whether some arbitrary parameters should be cooked up. Given this limited choice, I would opt for the first possibility. However, despite these climate models produce quantitative results, by and large, only qualitative conclusions should be derived and this always with a grain of salt.

In the following I present results obtained with the STACO model<sup>10</sup> as this allows me to draw on the largest set of simulations obtained for different treaty designs but for the same set of parameter values. Compared with most other climate models, STACO is richer in that the world is divided not only into 6 but 12 world regions. These include USA, Japan, European Union (EU-15), Other OECD countries (O-OECD), Central and Eastern European countries (EE), Former Soviet Union (FSU), Energy exporting countries (EEX), China, India, Dynamic Asian economies (DAE), Brazil and "Rest of the world" (ROW).<sup>11</sup> As with most coalition models, the (one-shot) decision about membership is based on discounted payoffs. However, in contrast to other climate models, STACO is simpler by restricting the emission path to be constant over time.<sup>12</sup> Nevertheless, all qualitative conclusions upon which I report below are confirmed by the results derived with the Climate World Simulation Model (CWSM) - a

---

<sup>10</sup> This model has been developed by Dellink et al. (2004). The subsequent results are compiled from Altamirano-Cabrera et al. (2006), Finus et al. (2004, 2005, 2006) and Weikard et al. (2006).

<sup>11</sup> EU-15 comprises the 15 countries of the European Union as of 1995. O-OECD includes among other countries Canada, Australia and New Zealand. EE includes for instance Hungary, Poland, and Czech Republic. EEX includes for example the Middle East Countries, Mexico, Venezuela and Indonesia. DAE comprises South Korea, Philippines, Thailand and Singapore. ROW includes for instance South Africa, Morocco and many countries in Latin America and Asia. The categorization follows Babiker et al. (2001).

<sup>12</sup> The bias introduced by this assumption is of minor importance as long as the decisions about membership is a one-shot decision and based on discounted payoffs. Note that in STACO overall magnitudes about accumulated emissions are calibrated such that they are in line with those of RICE.

fully-fledged CGE-model.<sup>13</sup> Moreover, in all papers mentioned under the STACO and CWSM in footnotes 10 and 13, respectively, sensitivity analyses have been conducted which confirmed all qualitative conclusions. This points to the robustness of qualitative conclusions.

## 3.2 Results

Results for the first set of alternative assumptions are displayed in Table 2. From this table, the following conclusions can be drawn.

In the presence of externalities, the potential gain from cooperation can be large. This is evident for instance by comparing global welfare in the global optimum with that in the Nash equilibrium which is three times larger.<sup>14</sup> But also partial cooperation can make a difference. For instance, a coalition of industrialized countries could raise global welfare by more than 50 percent compared to no cooperation which drops to 35 percent without the USA. It is also evident that not only the sheer number of participants is important for global welfare but also the identity of membership. For instance, a coalition of energy exporting countries (EEX) and China leads to a higher global welfare than a coalition between Former Soviet Union (FSU), Brazil and the Rest of the World (ROW). Generally, the importance of the identity for cooperation can be linked to two effects. Countries with a flat marginal abatement curve, like China, USA and India in this model, keep abatement costs low in a given coalition. Countries with high marginal benefits from global abatement, like EU-15, USA and Japan in this model, push for high coalitional abatement in a given coalition from which not only coalition members but also non-members benefit.

For this model, no coalition is stable for the base assumptions (see column 2). This changes with exclusive membership (see columns 4 and 6). Majority voting (unanimity voting) means that if an outsider to a coalition wants to join the coalition, this can be turned down if a majority of members (one member) is against accession. In the example, majority voting leads to one stable coalition (see column 4) and unanimity voting implies three stable coalitions (see column 6). This goes along with higher global welfare and higher global abatement. If we allow for multiple coalitions, then for unanimity voting the success of cooperation improves even more (see column 7).<sup>15</sup>

---

<sup>13</sup> See Eyckmans and Finus (2006a, b, 2007) where the issues a) single versus multiple coalitions, b) open versus exclusive membership and c) various transfer schemes, which I discuss subsequently, are also considered. Only the issue efficient abatement versus bargaining has not been considered with CWSM.

<sup>14</sup> In most CGE-models this difference is usually smaller in relative terms. See for instance Eyckmans/Finus (2006a) and Nordhaus/Yang (1996).

<sup>15</sup> The idea to consider multiple coalitions in the context of IEAs goes back to Carraro (2000). An analytical analysis of single versus multiple coalitions in the context of symmetric countries under various assumptions

For the understanding of these results, it is helpful to note the following relations. Given that regions in this model have a very heterogeneous cost-benefit structure, in the absence of transfers, an efficient allocation of abatement duties within a coalition leads to an asymmetric distribution of the gains from cooperation. This may imply that even some members are worse off than without cooperation. For instance, in most coalitions of which China is a member of a coalition, she has to contribute much to the joint abatement level (because of her flat marginal abatement cost curve), but benefits only little (because of low marginal benefits). In contrast, a region like the EU-15 contributes little to joint abatement (because of her steep marginal abatement cost curve), but benefits much (because of high marginal benefits). Thus, if at all, only coalition structures with coalitions that comprise of members with a similar cost-benefit structure can be stable. Otherwise, a disadvantaged coalition member would simply leave the coalition. That is, internal stability of a coalition would be violated. It is for this reason that regardless of the treaty design (open versus exclusive membership, single versus multiple coalitions), there is no stable coalition including the major industrialized regions, like USA, EU-15 and Japan, and at the same time non-industrialized regions, like China, India and ROW.

By the same token, advantaged countries have an incentive to join a coalition. That is, external stability would be violated. It is this last feature which explains why exclusive membership can make a difference compared to open membership. Not surprisingly, unanimity voting makes it easier than majority voting to keep advantaged countries out.

For a single agreement, the participation decision is only yes or no. Due to strong free-rider incentives, the decision is often no and hence only small and few coalitions are stable if at all. If multiple agreements are possible, even a country does not want to join the main crowd, it has a second option – forming another agreement with like-minded countries. For instance, the coalition of FSU, Brazil and ROW is stable for exclusive membership unanimity voting if only a single agreement can be formed (see column 6). If multiple coalitions are possible, other regions have an incentive to form their own agreement. Hence, the coalition structure

---

about membership rules is conducted in Carraro/Marchiori (2003), Finus (2003b) and Finus/Rundshagen (2003). These papers confirm the superiority of multiple compared to a single agreements as this is also done by Asheim et al. (2006) in a compliance model. A different conclusion is reached in Bosello et al. (2003) and Buchner/Carraro (2005b) with the FEEM-RICE model. However, this seems to be due to some bizarre assumptions about the abatement level and allocation chosen within coalitions. For instance, Bosello et al. (2003) report that no multiple coalition structure is conditionally rational (which they call weakly profitable). This means that the aggregate payoff of a group of countries *decreases* when they cooperate. This result is even more surprising because in a multiple coalition setting coalition members should benefit from the abatement activities of other coalitions.

with only a coalition of FSU, Brazil and ROW is no longer stable. However, five other coalition structures are stable including a coalition with these three regions.

Taken together, Table 2 allows for two major policy conclusions. Firstly, the open membership rule as applied in almost every international environmental treaty should not be taken for granted. Club good agreements (e.g., NATO, European Union and WTO) may not only enjoy a higher stability and success because the benefits from contributions are exclusive to members, but also because accession is limited through majority or unanimity voting. Put differently, the requirement to take decisions by consensus may not always be an obstacle for successful cooperation – a conclusion which will be confirmed in a slightly different context below.

Secondly, the effort to get as many countries as possible into one “climate boat” may not be the best strategy in the presence of free-rider incentives. Allowing for separate agreements among regions that have similar interests may foster the success of international agreements. In this light, the position of the USA only to ratify the Kyoto Protocol if developing countries would also join the protocol and take on climate responsibility is open to criticism.<sup>16</sup> However, their announcement in the aftermaths of their withdrawal that they would pursue their own climate policy cooperating with countries of similar interests may be viewed less critically. At least as an intermediate step, multiple agreements may be useful on the way to a global treaty as this is often the case with regional trade agreements that lead to more comprehensive agreements at later stages.

Table 3 illustrates the role of transfers. In order to separate effects, all other base assumptions are retained (i.e., single agreement, open membership). Generally, transfers can be implemented in many ways of which I report only on outcome-based rules.<sup>17</sup> That is, the surplus of cooperation within a coalition is distributed according to weights which sum up to one.<sup>18</sup>

---

<sup>16</sup> In 1997, the U.S. Senate unanimously passed the Byrd-Hagel resolution, which makes “meaningful” participation of developing countries a *conditio sine qua non* for ratification (The Byrd-Hagel Resolution, U.S. Senate, 12 June 1997, 105th Congress, 1st Session, Senate Resolution 98).

<sup>17</sup> In the literature (e.g., Rosen et al. 1998), transfers have been categorized into outcome-based, allocation-based and process based rules. The names refer to the main motivation for a transfer rule. For instance, outcome-based rules are earmarked to generate a particular outcome and hence the surplus from cooperation is shared in some way (see the subsequent footnote). In contrast, allocation-based rules focus on the initial allocation as this is for instance the case for the initial allocation of emission permits. Allocation-based rules are considered in Altamirano-Cabrera and Finus (2006) with STACO; process-based rules are analyzed in Bosello et al. (2003) with RICE.

<sup>18</sup> Let  $S$  denote an arbitrary coalition,  $S^N$  the singleton coalition structure corresponding to the Nash equilibrium and  $\pi_i(S)$  the payoff of country  $i$  without transfers if coalition  $S$  forms, then the payoff of a coalition member  $i \in S$  with transfer,  $\pi_i^T(S)$ , is given by  $\pi_i^T(S) = \pi_i(S^N) + \lambda_i [\sum_{j \in S} \pi_j(S) - \sum_{j \in S} \pi_j(S^N)]$  where the surplus is the term in brackets and  $\sum_{i \in S} \lambda_i = 1$ .

Among the many possible criteria according to which weights can be derived, I consider six schemes in Table 3.<sup>19</sup> The first two schemes may be called pragmatic (GDP and Emissions) whereas schemes 3 to 6 can be linked to some notion of equity (Equal Sharing, Ability to Pay, Inverse Emissions and Population). The pragmatic schemes more or less preserve the status quo and therefore have also been called sovereignty rules.<sup>20</sup> They favor industrialized countries with high current emissions and GDP. “Equal Sharing” simply splits the cooperative pie equally. “Ability to Pay” assumes weights inverse to the GDP/Population ratio. Thus, international pollution control serves as a vehicle of income redistribution. This scheme gives much weight to highly populated and/or poor regions. “Inverse emissions” appeals to the principle of historical responsibility for climate change and “Population” is based on the principle of “one man one vote”. Both schemes also favor developing countries and countries in transition as they have relatively low current emissions and a large population. The Nash Equilibrium and the Global Optimum are listed in Table 3 as benchmarks for the evaluation of the success of the various transfer schemes. In the case of multiple equilibria (as this is the case for the transfer scheme “Inverse Emissions”), the average payoff over all equilibria has been computed, which is the number besides the bracket in Table 3.

From Table 3 it is evident that the pragmatic schemes perform better than the morally motivated transfer schemes. Moreover, only for the pragmatic schemes a stable coalition emerges including industrialized and non-industrialized regions.<sup>21</sup> Additionally, we may recall from Table 2 that without transfers there is no stable coalition in this model for the base assumptions. However, also with transfers stable coalitions are relatively small and fall short of the global optimum. This leads to two conclusions.

Firstly, transfers are important to balance asymmetries and can improve upon the success of cooperation, though they cannot completely offset free-rider incentives. Hence, the establishment of international funds, like the Global Environmental Facility, in order to provide poorer countries with incentives to participate in IEAs (as this is the case for the Convention on Biological Diversity, the Montreal Protocol and the Kyoto Protocol) is an important step for successful international environmental cooperation.

Secondly, successful cooperation requires participation of industrialized and non-industrialized regions. This can only be achieved with moderate transfers which take into account the

---

<sup>19</sup> For a more detailed description see Altamirano-Cabrera and Finus (2006) and Weikard et al. (2006).

<sup>20</sup> In the context of allocation-based rules, the scheme “Emissions” would correspond to “grandfathering”.

<sup>21</sup> For allocation-based rules, as considered in Altamirano-Cabrera and Finus (2006) in the STACO-model, results are even more pronounced. Only for pragmatic rules stable coalitions exist; no coalition is stable under any morally motivated rule considered in this paper.

fundamental interests of all actors in climate change. The morally motivated transfer schemes would imply large transfers between industrialized and non-industrialized regions. Hence, no coalition of mixed membership is stable as long as only material payoffs are considered or a large redistribution of wealth is not seen as value in its own right (which is not the case in this model setting).

Before turning to the next point, it is worthwhile mentioning that I have only considered transfers within a coalition. The assumption implies that an outsider joining the coalition would automatically participate in the transfer schemes considered above. However, the assumption does not include at least two other interesting policy options.

The first option, though it may sound curious, is that an outsider pays another outsider for taking on more climate responsibility by joining the agreement. If the payment covers the additional abatement cost of the payee and provide the payer with a benefit in excess of his payment, this could be a viable option. For instance, this could mean that the USA, despite of not being and not becoming a member of the Kyoto Protocol could finance membership of China or India in order to join this treaty. In the CGE-model CWSM considered in Carraro et al. (2006) it is shown that such a strategy can work, implying a Pareto-improvement to all countries.

The second option is suggested by the current architecture of the Kyoto Protocol. The clean development mechanism (CDM) allows Annex-B-countries (i.e. mainly industrialized countries which accepted emission ceilings) to replace their own abatement activities by financing abatement activities in Non-Annex-B-countries (i.e. mainly developing countries without emission ceilings). Whether the CDM increases participation and the success of this protocol is at least from a pure game theoretic point of view not easy to predict. The reason is that in the absence of any strategic effects, CDM raises not only the welfare of signatories but also of non-signatories. If the latter effect is stronger, leaving the coalition becomes more attractive and CDM could have adverse effects. Therefore, it would certainly be interesting to evaluate these effects in an empirical climate model.

Until now, an efficient allocation of abatement within the coalition has been assumed. However, in reality, in most IEAs neither is the coalitional abatement level optimal nor is its allocation cost-effective. Typically, the coalitional abatement is below optimal levels<sup>22</sup> and

---

<sup>22</sup> This conclusion is supported for instance by empirical studies on the Montreal Protocol (Murdoch and Sandler 1997b), the Helsinki Protocol (Murdoch and Sandler 1997a), the Oslo Protocol (Finus and Tjøtta 2003), or the Kyoto Protocol (Böhringer and Vogt 2004).

abatement duties are specified as uniform emission reduction quotas.<sup>23</sup> In order to accommodate this observation and to provide a rationale for it, I compare the “Efficient Design” with three alternative treaty designs which are called “Common Quota”, “Median Quota Proposal” and “Lowest Quota Proposal” in Table 4.<sup>24</sup> (Again, the Nash Equilibrium and the Global Optimum are listed in Table 4 as benchmarks and average welfare is displayed in case of multiple coalitions). Common Quota departs from the standard assumption in one respect: coalitional abatement is chosen optimally, but not cost-effectively. That is, it is assumed that the coalition maximizes the aggregate payoff to this coalition as before but with the constraint that all members have to reduce their emissions by the same percentage from their non-cooperative levels (i.e. emission reduction quota). Median and Lowest Quota Proposal depart from the base assumption in an additional respect. The coalitional abatement level is not chosen optimally – it is not derived from joint but from individual welfare maximization. That is, each member makes a proposal for coalitional abatement which is optimal from its point of view, assuming that this is implemented in the form of a uniform emission reduction quota. If they agree on the median proposal, this corresponds to majority voting; if they agree on the lowest proposal, this corresponds to unanimity voting.

Since the Efficient Design corresponds to the base assumption we know from above that no coalition is stable. For the design Common Quota and Median Quota Proposal one coalition between India and ROW is stable. This coalition is also stable for the design Lowest Quota Proposal but global welfare is lower. However, there is an additional stable coalition of four regions with higher global welfare. The average global welfare for the Lowest Quota Proposal is higher than for the other two quota designs.

It is evident that in a given coalition and abstracting from stability considerations, any departure from the Efficient Design implies a global welfare loss. This welfare loss is particularly big in large coalitions and is particularly pronounced for the Lowest Quota Proposal.<sup>25</sup> However, including stability considerations, no coalition is stable for the Efficient Design without transfers. This is because this design implies an asymmetric distribution of abatement allocation

---

<sup>23</sup> The list of examples is long and includes the Helsinki Protocol, which suggested a 30 percent reduction of sulfur emissions from 1980 levels by 1993. Moreover, the "Protocol Concerning the Control of Emissions of Nitrogen Oxides or Their Transboundary Fluxes" signed in Sofia in 1988 called on countries to *uniformly* freeze their emissions at 1987 levels by 1995 and the "Protocol Concerning the Control of Emissions of Volatile Organic Compounds or Their Fluxes" signed in Geneva in 1991 required parties to reduce 1988 emissions by 30 percent by 1999.

<sup>24</sup> The idea of uniform emission reduction quotas goes back to Endres (1997) and Eyckmans (1997).

<sup>25</sup> For instance for the grand coalition, the Efficient Design implies a global welfare of 6,031 bln US\$ (as displayed in Table 4) whereas for the designs Common Quota, Median Quota Proposal and Lowest Quota Proposal we find 4589, 4565 and 3212 bln US\$ (not displayed in Table 4), respectively. See Altamirano-Cabrera et al. (2006), Table 1.

associated with an asymmetric distribution of the gains from cooperation. Moreover, it means that an ambitious coalitional abatement level is implemented. In contrast, the quota designs lead to a more symmetric distribution of the gains from cooperation and implement more modest coalition abatement targets which is pronounced particularly for the Lowest Quota Proposal. In the example, the departure from cost-effectiveness and ambitious abatement targets is rewarded. Because of strong free-rider incentives, the Lowest Quota Proposal is the most successful design.<sup>26</sup>

This conclusion is also confirmed if we allow for the possibility that quotas can be traded within a coalition as displayed in Table 5. Since trading constitutes a win-win-situation for all coalition members, coalitional members' payoffs increase and hence also global welfare for a given coalition structure.<sup>27</sup> Hence, the difference between the Efficient Design and the three quota designs is only due to the "modesty effect", i.e. lower coalitional abatement levels in a given coalition since trading leads to cost-effectiveness.<sup>28</sup>

The Pareto-improving effect of trading has two implications. Firstly, it is less attractive to leave a coalition. That is, internal stability is strengthened. Hence, all coalitions that have been internally stable without trading are also internally stable with trading and may be some additional coalitions. This is why for instance the coalition between India and ROW which is stable in Table 4 is also internally stable in Table 5. Secondly, it becomes more attractive for outsiders to join a coalition. That is, external stability is weakened. This is why for instance the coalition between India and ROW is no longer externally stable with trading. However, this external instability of small coalitions is rewarded by larger stable coalitions in this example. For all three quota designs, average global welfare of stable coalitions with trading is higher than without trading. The ranking within the three quota designs according to average global welfare remains the same: the highest average global welfare is obtained for the design Lowest Quota Proposal, followed by Median Quota Proposal and then Common Quota.

From Table 4 and 5 the following three policy conclusions emerge, taking in consideration that IEAs have to operate in a second-best world, i.e. participants face free-rider incentives but no third party can enforce a treaty. Firstly, in the absence of transfers, there is a rationale for implementing uniform emission reduction quotas. Despite being not cost-effective, quota

---

<sup>26</sup> A similar conclusion is reached in a compliance model in Finus/Rundshagen (1998).

<sup>27</sup> For the grand coalition (as reported in the previous footnote), global welfare for the designs Common Quota, Median Quota Proposal and Lowest Quota Proposal is now given by 5547, 5703 and 3345 bln US\$ (not displayed in Table 5), respectively. See Altamirano-Cabrera et al. (2006), Table 1.

<sup>28</sup> That modesty can pay is also confirmed in theoretical models with symmetric players. In a membership model this is shown in Finus (2004) and in a compliance model in Barrett (2002).

agreements may be more successful as the gains from cooperation are more equally shared. Secondly, the pressure for consensus treaties is not always an obstacle for successful cooperation. If modest abatement targets are compensated by a sufficiently high participation, the overall effect may well be positive. This conclusion (“better a bird in hand than two in the bushes”) gains momentum when considering (by taking a wider perspective) that many IEAs have been established first with modest abatement targets, as for instance the Montreal Protocol on the Reduction of CFCs. Subsequently, amendment protocols followed that gradually tightened abatement targets and broadened the base of regulated pollutants.<sup>29</sup> Thirdly, emission trading, as for instance established for the first time within the Kyoto Protocol, can increase participation and the success of a treaty. This stresses the importance of this instrument not only for cost-effectiveness but also for the incentive of participation in IEAs.

#### **4. Possible Extensions**

Some of the results in the previous section have a positive dimension. With positive dimension I mean that these results help to rationalize apparently second- or third-best designs of actual IEAs. Examples include the following conclusions: a) uniform may be superior to differentiated and cost-effective emission reductions; b) trading of uniform emission reduction quotas has a positive effect on participation and efficacy of IEAs, c) modest may be superior to ambitious emission reductions. Of course, at the same time, these insights provide also some normative guidance for the pragmatic design of future IEAs.

However, some results seem to have an exclusive normative dimension. For instance, we concluded that d) multiple are superior to single agreements and e) exclusive is superior to open membership. Clearly, all current IEAs are single agreements as there is for instance only one Montreal Protocol and one Kyoto Protocol. Of course, one may argue that recent developments seem to indicate that a group of countries around the USA may set up an additional climate treaty. However, without any doubt, all current environmental agreements of global scale do not restrict accession and therefore adhere to the principle of open membership. Moreover, membership in Table 2 to 5 under various institutional assumptions is neither in line with participation in the current Kyoto Protocol nor with the expectations of probably most experts on climate change. For instance, the “current Kyoto coalition”, corresponding roughly to a coalition of Japan, European Union (EU-15), Other OECD countries (O-OECD), Central and Eastern European Countries (EE) and the Former Soviet Union (FSU) according to the regional classification in STACO is not stable under any of the assumptions considered

---

<sup>29</sup> See for instance the Appendix in Finus (2007).

here. In contrast, depending on the assumptions, a coalition of the Former Soviet Union (FSU), Brazil and Rest of the World (ROW) may form a stable coalition (e.g. Table 2).

A partial answer to resolve this puzzle could be to point out that it is very likely that the Kyoto coalition will indeed turn out to be unstable as predictions suggest that many participants will not be able to fulfill their abatement obligations in the first commitment period 2008-12. It may also be pointed out that politicians are not completely rational, pursue other goals than maximizing the welfare of its citizens or simply do not know for instance that exclusive is superior to open membership. After all, economists have put forward numerous normative policy recommendations - not only in environmental economics - that have not fallen on fertile ground. It is certainly beyond the scope of this article to provide a complete solution to this puzzle. However, it is clear that a stylized model can only capture a small amount of the diverse factors that play a role in the actual setting under which the Kyoto Protocol is operating. Some of these factors may not even be known. Moreover, even from a purely theoretical point of view, it is obvious that some aspects of coalition formation have been neglected in the previous analysis. In the following, with reference to Table 1, I will discuss some aspects which I believe are the most prompting one to be considered in future research.<sup>30</sup> It will become apparent some of the issues have already been treated, though some under very specific assumptions with no general and clear-cut conclusions.

### *1) Sequential Participation Decisions*

The base assumption implies that all countries decide simultaneously upon their participation. However, casual empirical evidence suggests that signature and ratification of environmental treaties occur sequentially. Typically, some initiators kick off the negotiation process, a core group of countries participate in an agreement and other countries may follow suite at a later stage.<sup>31</sup>

However, modeling coalition formation as a sequential process introduces a couple of conceptual problems that are difficult to solve. For instance, in the case of heterogeneous countries, the sequence in which countries decide upon their participation has to be endoge-

---

<sup>30</sup> I will not cover at least one issue that I believe is also important in the context of IEAs. This is the link between cooperation on environmental issues and R&D. The analysis of the link between environmental and some other policy issues in order to improve the prospects of IEAs has been initiated by Folmer et al. (1993) and Folmer/van Mouche (1994) in the context of compliance models. Later contributions analyzed issue linkage also in the context of membership models as for instance Barrett (1997) and Carraro/Siniscalco (1997). Recently, this topic has gained new momentum with the focus on how the design of an IEA affects endogenous technological progress and how this affects the success of an IEA. See for instance Golombek/Hoel (2005).

<sup>31</sup> The evolutionary development of some IEAs is reported for instance in the appendix in Finus (2007).

nized. In a model with a dynamic payoff structure, which means in particular that the payoff at time  $t$  depends on accumulated emissions (i.e. stock of emissions or concentration in the context of greenhouse gases), the number of strategic options increases dramatically. Optimal abatement levels as well as optimal participation strategies have to be determined at each point in time in a recursive way.

Until now only a few preliminary results are available with limited policy-relevance. The reason is that all results have been derived from highly stylized models with symmetric countries<sup>32</sup> and many other very specific assumptions.

The simplest extension assumes a sequential coalition formation process according to the sequentially move unanimity game of Bloch (1995) as considered for instance in Finus/Rundshagen (2006b) and Ray/Vohra (2001). In this game, membership is exclusive and multiple coalitions may form. Once a coalition has formed by unanimous voting, membership cannot be revised.

From these papers it appears that time introduces an additional strategic component: countries have an incentive to commit to no or low cooperation in the early stage of the formation process, thereby increasing the pressure on other countries for cooperation. Moreover, due to strategic considerations, full cooperation may not be possible; the equilibrium coalition structure may not even be Pareto-optimal. Surely, this is hardly a surprising result and not so much different from what has been obtained with simpler models assuming a simultaneous participation process. Moreover, contrary to reality, the initiators are not those countries that establish an agreement but those which free-ride.

A more interesting analysis is conducted in Rubio and Ulph (2002, 2007). Both papers consider in the base model of an open membership single coalition game a stock pollutant and the decision about membership over time. Models are simple as countries have only a dichotomous choice between no abatement and abatement. In the first paper, only two periods are considered. In line with real world observations, the authors show that participation rises in the second period compared to the first period. However, this seems to be an artificial result as the authors admit because the finite time horizon introduces an end point bias. In the last period (here the second period), no environmental damages in subsequent periods have to be considered. Hence, optimal abatement is low and participation is associated with low abate-

---

<sup>32</sup> Symmetric players means that all players have the same payoff function. Sometimes this is also called ex-ante symmetric players as ex-post signatories and non-signatories receive different payoffs since they chose different abatement levels.

ment costs. This qualification is confirmed by the infinite horizon version in the second paper. Here the authors find that participation is a non-increasing function of time.

Hence, it appears that only two policy-relevant conclusions can be derived from these papers. Firstly, under reasonable assumptions the equilibrium stock of emissions increases over time, approaching some steady state. However, this is already known from empirical climate models like RICE (Nordhaus/Yang 1996) and WIAGEM (Kempfert 2005) though the particular focus on partial instead of full cooperation is new. Secondly, the authors confirm a conclusion for the assumption of efficient abatement within a coalition that not only in a static but also in a dynamic setting the success of coalition formulation is negatively related to the level of environmental damages.<sup>33</sup> However, despite these critical remarks, no doubt, the research direction kicked-off by Rubio/Ulph (2002 and 2007) is an important and promising one if it is combined for instance with some ingredients I mention below. This may allow deriving more realistic conclusions.

## 2) *Sequential Abatement Decisions*

The standard assumption implies that the coalition, acting as a single player, and the outsiders, acting as singletons, simultaneously chose their optimal abatement policy. An alternative assumption, introduced by Barrett (1994) and further developed by Diamantoudi/Sartzetakis (2006) and Rubio/Ulph (2006), is that this happens sequentially. He assumes that the coalition acts as a Stackelberg leader, taking the (aggregate) best response of outsiders in consideration, when choosing its abatement. This informational advantage can reduce the degree of leakage and leads to larger and more successful coalitions.<sup>34</sup>

Until now, this result has only been established for the assumption of symmetric players and three specific payoff functions.<sup>35</sup> Nevertheless, it seems suggestive that this result should hold more generally. From a theoretical point of view, however, there are two problems with this alternative assumption. Firstly, in a model with symmetric players and/or payoff function comprising only cost and benefits from abatement, the informational asymmetry does not follow from the model itself and requires some exogenous motivation. Secondly, in the special context of coalition formation, the question arises why potential coalition members should assume that they lose their informational advantage once they leave the coalition.

---

<sup>33</sup> This result is summarized and explained in Finus (2003a) for the static setting. See in particular Result 2 in section 3.3.2.

<sup>34</sup> Leakage effect means that non-signatories lower their abatement as a reaction to higher abatement of signatories. For instance, if the members of a climate agreement reduce their energy demand, energy prices drop, inducing higher demand by outsiders.

<sup>35</sup> For an overview see Finus (2003a), Table 3.2, p. 108.

From an applied point of view, the result is more interesting, suggesting that successful cooperation should take in consideration the strategic reaction of outsiders.

### *3) Subjective Payoffs*

STACO like other climate models group countries into regions for tractability. Even though in climate negotiations some countries coordinate their strategies, like the European Union or the group of G77 countries, after all, countries take sovereign decision when it comes to accepting final abatement targets and signing and ratifying a treaty. This introduces a bias in the incentive structure. Anything else being equal, we should expect that this aggregation into regions overestimates the success of cooperation derived from climate models. This qualification calls for a model with more disaggregated data. However, apart from the fact that more disaggregate data is currently not available, such an extension would require a substantial amount of additional computer resources.<sup>36</sup>

Moreover, as far as I am aware, all climate models derive the parameters of their abatement cost functions and benefit functions from some “scientific sources”. Apart from the difficulty to monetize environmental damages in general, these estimates try to evaluate “objective” damages. For instance, most climate models assume high environmental damages in developing countries summarized in the region “Rest of the World” (ROW). From a normative view point this may not be seen as a problem. However for a positive analysis, this may introduce some bias. There are some reasons to believe that the subjective perception of damage costs of ROW is substantially lower than those assumed in most climate models, overestimating the interest of this heterogeneous region for cooperation.

One possibility to improve upon estimates for a positive analysis could be to use data from contingent valuation studies to calibrate climate models. For instance, Böhringer and Vogt (2004) suggest based on a survey of various studies that the willingness to pay for global greenhouse gases is even in industrialized countries rather low. However, it is also evident from their survey that the current data basis is very small - only few country studies are currently available.

### *4) Political Process*

The base model assumed that abatement within a coalition is derived from joined welfare maximization of coalition members. As already pointed out, there are at least three problems with this assumption. Firstly, it is not consistent with the notion of non-cooperative game

---

<sup>36</sup> In the context of only 12 world regions, there are already more than four billion different coalition structures in the case of multiple coalitions. This number increases exponentially with the number of regions.

theory. Secondly, it certainly overestimates the sense of cooperation among governments, even though they are willing to take joint action. Thirdly, already casual empirical evidence suggests that this simply does not reflect what is really happening in international negotiations: coalitional abatement is neither optimal nor is its allocation cost-effective. Certainly, an improvement to accommodate this critique was the previous assumption of a bargaining process over uniform emission reduction quotas. However, this can only be seen as a first step.

Firstly, the set of bargaining options was restricted (e.g. Median Quota and Lowest Quota Proposal), may not apply to all situations of international environmental negotiations and the choice of the negotiation regime was not endogenized. Moreover, in the case of transfers, though several options were considered, the choice of the transfers scheme (i.e. the choice of weights) was not endogenized. Secondly, and most important, the political process that defines the positions of governments in international negotiations has not been considered explicitly (e.g. Michaelowa 1998). This shortcoming is shared by all coalition models as far as I am aware and was noted already by Carraro/Siniscalco (1998). Hence, one possibility to improve upon the predictive power of coalition models could be to include public choice aspects, like interest groups, governmental decision processes and parliamentary voting. However, apart from increasing the complexity of models, it should be expected that it will be difficult to identify and quantify political variables for an empirical analysis.<sup>37</sup>

##### *5) Non-material Payoffs*

The standard assumption implies that decisions about participation and abatement are solely based on costs and benefits from abatement (i.e. material payoffs). From a normative as well as from a positive perspective one may argue that this view is too narrow. From a normative perspective one may argue that it is exactly for this reason why cooperation proves so difficult in reality and why game theoretic models confirm this pessimistic view. From positive perspective there may be the conjecture that also aspects like reputation or ethical considerations of fairness play some role when negotiating an IEA and deciding on participation and compliance. Moreover, one may presume that these aspects are at least partially responsible for explaining some of the success stories of IEAs, like for instance the frequently cited Montreal Protocol.

---

<sup>37</sup> As far as I am aware, only Bucholz et al. (2005) and Altamirano-Cabrera (2007), chapters 7 and 8, consider public choice aspects in transboundary pollution. However, Bucholz et al. (2005) is a highly stylized model with only two symmetric players. Hence, the problem of coalition formation does not arise. Altamirano-Cabrera (2007) is a very promising approach as he considers public choice aspects and coalition formation in the STACO-model. However, the analysis must be judged as preliminary as it is flawed by some inconsistent assumptions.

Though the normative perspective is an extremely relevant and important issue of non-material payoffs, scientifically it is of limited interest. Even without resorting to non-material payoffs, full cooperation could easily be achieved in theory if all actors behaved differently. If all actors would not only consider their own payoff but also the benefits occurring to others from their abatement when a) choosing their abatement level and b) deciding whether to join or leave an IEA, the grand coalition, implementing the social optimum, would be the unique stable outcome. Even in a richer model, the normative dimension of the problem would call for the complete internalization of all transboundary externalities across all actors. That is, all actors care also for the well-being of others, regardless how this is defined. A more interesting issue is certainly the question: How is it possible to change the preference of actors so that they consider externalities they impose on others? This question is related to the role and formation of norms and the institutions that help in shaping them. Unfortunately, as far as I am aware, no such research in the context of IEAs has been conducted.

The positive perspective is scientifically interesting but also difficult. The main question that arises is to which extent and in which sense do ethical aspects guide the behavior of actors? However to the best of my knowledge, this question has not really been addressed yet.<sup>38</sup> On the one hand, there is a couple of theoretical papers which have analyzed how non-material payoffs influence the success of IEAs. On the other hand, there is one empirical paper that conducted a questionnaire among employees in governmental and international institutions involved in climate change on the importance they attach to different notions of equity. Though this survey is important, by its nature we should expect a bias between stated preferences and those underlying actual negotiations. In the following, I briefly review these papers.

Jeppesen/Andersen (1998) and Hoel/Schneider (1997) add a term to the "classical payoff function", which the former call a "non-material payoff", representing the idea of fairness, and the latter "non-environmental cost of breaking the agreement". They assume that all countries' (symmetric) payoffs comprise benefit minus cost from abatement (material payoff) but that non-participants receive additionally disutility from being outsiders (non-material payoff) where disutility increases with the number of countries participating in an IEA. Not surprisingly, both papers find that the higher the disutility of being an outsider, the larger will be the

---

<sup>38</sup> There are only few empirical studies on the behavior of countries in the context of IEAs. In the context of climate change, see Böhringer/Vogt (2004); in context of sulfur emissions, see Murdoch/Sandler (1997a) and Finus/Tjøtta (2003) and in the context of CFCs, see Murdoch/Sandler (1997b) and Swanson/Mason (2003). However, none of these papers include ethical aspects.

equilibrium coalition. At least technically, this result seems obvious, as it follows almost by assumption.<sup>39</sup>

Lange/Vogt (2003) assume that the utility of a country is the weighted sum of the absolute material payoff function and the relative material payoff function. The relative material payoff function obtains its maximum for the average payoff of all countries. Any deviation from the arithmetic mean is interpreted as inequality and counts as disutility. For the assumption of symmetric countries, they show that if relative payoffs receive a sufficiently high weight compared to absolute payoffs, the grand coalition emerges as an equilibrium coalition. Unfortunately, they do not analyze which equilibrium emerges if relative payoffs receive some weight but not sufficiently much in order to stabilize the grand coalition. Instead, they run simulations for a particular benefit and cost function assuming that there are 12 countries in total of which half care exclusively about absolute payoffs (A-countries) and half of them care exclusively about relative payoffs (B-countries). Their example illustrates that there can be multiple equilibria, comprising a different mix of A- and B-countries. However, given that simulations are only based on one parameter set, no general conclusions can be drawn from this example.

A similar approach is pursued in Peters/Schuler (2006). However, there are two differences. Firstly, disutility does not stem from payoff differences but from abatement differences. This assumption has some connection to the assumption of uniform emission quotas discussed in section 3. However, now the preference for uniformity is explicitly captured in the payoff function which may be judged as an advantage. Secondly, inequality is not simply measured as the deviation from the mean but by the variance. This corresponds to a wider definition of inequality. Even if a country chooses an average abatement level, this may be associated with negative utility as long as this average deviates from the maximum and minimum abatement level. Basically, Peters and Schuler confirm Lange/Vogt (2003) that if inequality aversion plays a sufficiently important role in countries payoff function, the grand coalition will emerge as a stable agreement. If inequality aversion is below this threshold, they show that multiple equilibria may emerge that range from no cooperation to partial cooperation with almost full participation. Moreover, they demonstrate that no correlation between inequality aversion and the size of coalitions can be established.<sup>40</sup>

---

<sup>39</sup> A non-obvious result in Hoel/Schneider (1997) is derived from an extension which considers the role of transfers paid by members to non-members for increasing their abatement efforts. However, their conclusion that transfers may have a negative impact on cooperation is flawed by their conceptual setting as I argue in Finus (2003a), section 3.3.4.

<sup>40</sup> A similar inconclusive result is obtained in Lange (2006). Though this model is very rich in terms of the policy setting, the analysis is incomplete as it does not consider external stability.

Apart from the difficulty to reach general conclusions about the role of inequality aversion in the models by Lange/Vogt (2003) and Peters/Schuler (2006), there are at least two more problems. Firstly, there is no empirical evidence that would back either one of two different measures of inequality aversion or suggest another one. For instance, in the context of global warming, fairness has been associated with the call for different responsibilities. This may be in line with similar abatement obligations within the group of developed countries, countries in transition and developing countries but certainly not across groups. Of course, the models assume symmetric countries, but even in such a setting one could imagine that the herd instinct as it has been called by Peters/Schuler (2006) is only felt within the group of coalition members and non-members but not among both groups. Secondly, neither of the two papers offers conclusions with respect to global welfare and global abatement. Though in a symmetric setting one may conjecture that there is a correlation between the size of stable coalitions and global welfare as well as abatement, this has to be analyzed.

Finally, I turn to the questionnaire by Lange et al. (2007). They contacted 1695 people (of which 230 responded) working for governmental bodies and international organizations involved in climate change policy. Questions about the importance of six equity rules for short- and long-term climate policy were put forward. Responses served as an input into various econometric tests. They find that equity criteria are generally considered as important by the respondents, though this was more important for respondents from developing countries and countries in transition than for those from industrialized countries. The polluter pays principle (i.e. equal ratio between abatement costs and emissions) and the poor losers principle (exemptions from abatement duties due to very low GDP) are the most widely accepted equity principles. Moreover and most important, there is a strong correlation between the importance attached to a particular equity criteria and the economic impact associated with it. That is, respondents favored those equity criteria that are associated with relatively low abatement obligations for their countries of origin.

Overall, this seems to suggest that equity criteria are just disguised means in the political arena in order to push through strategic economic goals in negotiations. Whether this negative conclusion holds generally will have to be investigated in future research. In particular, one may expect that there is a bias between stated and actual preferences. Moreover, stated preference for equity criterion of respondents may differ from those which actually entered into agreements. Therefore, it would be interesting to investigate not only stated but revealed preferences by investigating the correlation between the design of actual IEAs and their implied equity principle. For instance, the equity criteria “sovereignty rule” in Lange et al.

(2007) seem to have played an important role in the past as it is defined as equal percentage reduction of current emissions, which I referred to as uniform emission reduction quotas in section 3. See in particular footnote 23.

#### 6) *Uncertainty*

The standard assumption implies that all payoff functions are known with certainty. This is surely a heroic assumption for several reasons. For instance, the impacts of climate change are highly uncertain. This concerns the impact of greenhouse gas concentration on the climate, the impact of the climate on the environment and the relation between changes of environmental quality and its evaluation by people. In particular, the large time lag between emissions today and future environmental damages due to the accumulation of emissions makes estimates difficult. However, also the prediction about future abatement costs are difficult as future abatement technologies, economic growth and the relation between growth and emissions are not known today. This suggests that there is uncertainty regarding the parameters of the payoff function and its exact functional form. Moreover, we may suspect that there is also asymmetric information (e.g. Caparrós et al. 2004). Though representative of countries may have some idea about their own evaluation of the climate problem and possible solutions, they face uncertainty about the perception of their neighbors.

In order to study the strategic implications of uncertainty, it is necessary to specify how agents form expectations. This requires making assumptions about the distribution of the random parameters and the functional forms. Moreover, it requires specifying whether and how uncertainty is resolved through learning. In reality, learning is usually imperfect, meaning that the true state of the world is only revealed gradually but never complete. Methodological, this would mean that beliefs are sequentially updated as more information becomes available (e.g. through Bayesian learning). However, this would lead to very complex models. Therefore, as far as I am aware, all models on uncertainty and learning assume perfect learning. That is, if information is revealed, the true state of the world is revealed to all countries. Despite this simplification, clear-cut conclusions are difficult to obtain as there are basically two opposing effects.

On the one hand, the possibility that better information will be available in the future which reduces uncertainty leads to the suggestion that abatement activities should be delayed. For instance, it may turn out that cheaper abatement options will be available in the future. Hence, current abatement activities may turn out to be unnecessarily costly. In particular, if abatement today is associated with high fixed costs, this can make it more difficult to switch to

cheaper options in the future (Kolstad 1996). On the other hand, the irreversibility of greenhouse gas accumulation calls for early action in case it turns out that climate damages are more severe than expected (Ulph 1998). Moreover, current stricter environmental regulation may boost technological progress, making abatement cheaper in the future.

However, even in the simple models of Ulph (1998), Ulph/Ulph (1997) and Ulph/Maddison (1997) it is difficult to obtain general conclusions. These models consider only two periods with a stock externality, a fixed abatement technology, only uncertainty with respect to damages, which can only be high or low (e.g. there are only two states in the world), and two symmetric countries which may only differ in damage costs. In the case of no learning, abatement decisions are taken on expected damages. For their model assumptions, this case corresponds to the case of no uncertainty, except that the value of the certain damage parameter is replaced by its mean value.<sup>41</sup> In the case of learning, true damages are revealed after the first period and countries can take optimal decisions.

These papers confirm that total accumulated emissions are higher in the Nash equilibrium than in the social optimum, with and without learning. They also show that the value of learning (i.e. in the form of higher global welfare) is usually positive in the social optimum, except for the special case of negatively correlated damages in which case this value is zero. This is of course not surprising since more information should lead to better decisions. More surprising are two findings.

Firstly, only for specific payoff functions is it possible to conclude whether first period abatement levels are higher or lower with than without learning. This holds for the Nash equilibrium and the social optimum. This suggests that no general conclusions are possible whether early action is sensible. This qualification gains even more momentum when considering that countries are usually asymmetric, damage costs parameters may be distributed over a large interval and future abatement costs are also uncertain. Secondly, if damages are sufficiently negatively correlated in these models, the value of information can become negative in the Nash equilibrium. That is, in a strategic setting, learning can be bad. The intuition is that without learning, both countries are ex-ante and ex-post symmetric and hence choose symmetric abatement levels in both periods. Though abatement is below socially optimal levels, it is conducted cost-effectively as both countries have the same abatement cost function. With learning, countries choose different abatement levels as they find out ex-post that their true damage costs differ. This implies a further departure from the social optimum.

---

<sup>41</sup> They assume - as all other models do as far I am aware - that utility is linear in payoffs, implying risk neutrality, and hence the theorem of certainty equivalence holds. See footnote 45.

The driving force of this negative conclusion about the value of learning in a strategic setting is also present in a simple static three country model with coalition formation in Na/Shin (1998). Countries have the same payoff function which differs only in the damage parameter. This parameter is uniformly distributed with the same expected value for all countries. Without learning, which they call ex-ante negotiations, no information is disclosed; neither before the first stage nor before the second stage of coalition formation. Hence, the expected payoff from forming a coalition is the same for all countries. Hence, the grand coalition is stable in their model. With full learning, which they call ex-post negotiations, the true damage parameter is revealed to all countries before the first stage. For the base assumption of joint welfare maximization of the coalition, this implies symmetric abatement levels within a coalition from which countries benefit differently due to different damages. Hence, as no transfers are assumed, the grand coalition is not stable anymore in their model.<sup>42</sup>

It is evident that the conclusion on the negative value of information needs some caution as all the models mentioned above make very specific assumptions. This is also true for the coalition models by Kolstad (2007) and Kolstad/Ulph (2006). Both papers consider the case of a positive correlation and no correlation of damage parameters between an arbitrary number of countries. Contrary to the papers above, they also find a negative value of information in the case of a positive correlation of uncertainty. However, their payoff function is very special: damages and abatement costs are linear and hence optimal abatement decisions are boundary solutions, meaning that either abatement or no abatement is optimal. Moreover, in the case of no correlation of uncertainty, it turns out in these models that the value of information may become positive.<sup>43</sup> Such mixed evidence also holds for the case of partial learning which means that uncertainty is resolved after the first stage but before the second stage. Thus, only membership decisions have to be taken without knowing the true damage parameters.<sup>44</sup>

Taken together, we may conclude that uncertainty is a very important and relevant aspect of climate change. However, our understanding what it means for the strategic interaction between countries and for the prospects of cooperation is still very basic and incomplete. The

---

<sup>42</sup> Fujita (2004) claims that this conclusion would disappear with transfers and more than three countries. However, he wrongly claims that his stability concept of the CS-core would be identical to internal and external stability as applied in Na/Shin (1998). Thus, results cannot be compared.

<sup>43</sup> Also in the STACO-model with uncorrelated uncertainty of abatement cost and benefit parameters (and continues choices of abatement) it is shown in Dellink et al. (2005) that in the case of full learning the value of information can be positive in the context of coalition formation.

<sup>44</sup> The strong dependency on assumptions is also evident from Ulph (2004). This model is in the tradition of the models of Rubio and Ulph (2002, 2007) on dynamic coalition formation. Despite the assumption of symmetric players and only a dichotomous choice of abatement levels (due to linear abatement and damage cost functions) a comparison of no learning and full learning has to resort to simulations with mixed evidence.

main problem is the complexity of uncertainty and learning which makes it necessary to work with very simple models which do not allow drawing general conclusions. Most models are static or consider only two periods, assume symmetric countries and consider only uncertainty of climate damages. Moreover, none of the models on coalition formation considers the possibility that countries can take actions to reduce uncertainty. For instance, apart from the possibility to invest in research and development in order to reduce scientific uncertainty, there may be the possibility that increased abatement efforts reduce the spread of uncertain damages. If decision-makers are sufficiently risk averse, this could provide an incentive for more cooperation.<sup>45</sup>

### *7) Coordination*

The claim that IEAs have to be self-enforcing rests on the observation that there is no supra-national authority with enforcement power. Though I agree with this view, this neglects at least two issues. From a positive point of view the question arises why governments resort to international law when they have disputes about the interpretation of treaties or accuse other parties for not meeting their obligations? Moreover, there are international organizations like UNEP, the World Bank and the WTO that are involved in transboundary environmental externalities in one or another way. For instance, the WTO is involved in settling trade disputes also related to environmental standards. UNEP coordinates some environmental policies and calls for instance for increasing efforts to combat global warming. The World Bank manages some projects related to the CDM and the Global Environmental Facility. Ignoring such phenomena in models means simply that no rationale can be provided for the existence of these institutions.

From a normative point of view the question arises whether these or other international institutions, even though they cannot enforce IEAs, cannot play a constructive role as coordinators or mediators. This concerns for instance the stage of negotiation, signature, ratification and implementation of an IEA. Also international law may at least provide some guidance of good conduct. Again, ignoring such phenomena may mean to ignore relevant instruments for increasing international cooperation in mitigating transboundary environmental problems.

Taken together, I would strongly advocate supplementing game theoretic analyses by insights from international law, institutional economics and political sciences, to name only three disciplines, respectively methods. That this can be a successful strategy shall be briefly illus-

---

<sup>45</sup> Such an idea is pursued in Endres/Ohl (2000, 2003). They show in a simple static prisoners' dilemma game that if agents are sufficiently risk averse and cooperation can reduce the spread of uncertainty of payoffs, partial or full cooperation can be an equilibrium.

trated with two examples. Though both examples remain in the domain of game theory, they provide at least some hint how such a research strategy could look like.

In Finus/Rundshagen (2006b) it is shown that in the case of multiple stable coalitions a third party can bring about a Pareto-improvement to all countries. This third party can help to coordinate participation strategies, despite it has not enforcement power.<sup>46</sup> Though this model is very simple, it is easy to imagine that a third party may also be successful in coordinating abatement strategies. Moreover, in richer models, coordination may be extended to many other aspects of the design of IEAs, like transfer payments, monitoring and dispute settlement.

Another interesting aspect of coordination is analyzed in Carraro et al. (2004), Haeringer/Courtois (2005), Rubio/Casino (2005) and Rutz (2002), chapter 6. Though these models differ in detail, they all consider the effect of a minimum participation clause. The motivation derives from the fact that almost all IEAs concluded in the past operate under such a clause. For instance, Rutz (2002) reports that only 2 out of 122 IEAs that he considered had no such rule. In juridical terms, a minimum participation clause specifies the number instruments of ratification, acceptance, approval or accession that have to be deposited before a treaty enters into force. In the game theoretic jargon, all instruments simply mean becoming a signatory and “enters into force” means that only from now onwards agreed obligations are considered as binding for all members to an IEA. For instance, the Kyoto Protocol required a double-trigger before it entered into force. Firstly, at least 55 parties had to ratify the treaty by their national parliaments. Secondly, parties had to account for at least 55 % of the total 1990 CO<sub>2</sub>-emissions of all Annex-B parties.

All papers mentioned above analyze this feature in the base model with symmetric countries in a two step procedure. Firstly, they show that a stable coalition is small without this clause. This is because if the number of signatories increases sufficiently, it becomes attractive to leave a coalition, given that all other countries remain in the coalition. More precisely, if we let  $n$  denote the size of a coalition, then coalitions for which  $n^* \geq n \geq 1$  is true are internally stable and those for which  $n > n^*$  is true are not internally stable. Since for symmetric countries it holds that if  $n$  ( $n+1$ ) is an internally stable (internally unstable) coalition, then  $n-1$  ( $n$ ) is an externally unstable (externally stable) coalition, we can state: coalitions for which  $n^* > n \geq 1$  are externally unstable and those for which  $n \geq n^*$  are externally stable. Taken together, coalition  $n^*$  is internally and externally stable and hence stable.

---

<sup>46</sup> A similar idea is analyzed in simple matrix games in Ecchia/Mariotti (1998).

Secondly, it is shown that if a minimum participation clause  $n^\#$  is set above the equilibrium coalition size  $n^*$  without such a clause, e.g.  $n^\# \geq n^*$ , the coalition with  $n = n^\#$  members is stable. The argument runs as follows. If the number of signatories is  $n = n^\#$ , then this is a stable coalition since leaving the coalition meant that the entire coalition would break apart. This does not pay as the payoff of a signatory with  $n = n^\#$  members is higher than in the Nash equilibrium with  $n = 1$ . Since any coalition  $n \geq n^*$  is externally stable,  $n = n^\#$  is stable. It is also argued that a coalition with  $n > n^\# \geq n^*$  members cannot be stable because it is not internally stable and the threat of a complete resolution would be lost. Of course, setting the minimum participation strictly below  $n^*$  cannot be an equilibrium as a coalition for which  $n < n^*$  is true is not externally stable.

All papers conclude that a minimum participation clause works like a coordination device. It improves upon the prospects of cooperation because global welfare increases with the coalition size and  $n^\# \geq n^*$ . Moreover, in Carraro et al. (2004) the choice of the minimum participation clause  $n^\#$  is endogenized. This basically means to show that all countries are ex-post (without knowing whether they will be a signatory or non-signatory) better off with proposing  $n^{\#*}$  than any other proposal  $n^\# \neq n^{\#*}$ . They conclude that for the most common payoff functions countries will agree to set the minimum participation  $n^{\#*}$  to full participation.

I go along with this conclusion and also think that these contributions help to explain an institutional feature of actual treaty-making. However, I have three reservations.

Firstly, the question arises whether it is credible to assume that a member in a coalition with  $n = n^\# > n^*$  members should expect that leaving the coalition causes the entire coalition to break up into singletons. After all, all papers conclude that a coalition with  $n^*$  members is profitable and stable without clause. I also doubt whether a coalition with  $n > n^\#$  members cannot be stable. If countries are not myopic, then they know that a deviation may trigger further deviations. Hence, if the payoff of a signatory at  $n > n^\#$  was larger than as a non-signatory at  $n = n^\#$  deviation would not pay. Hence, a coalition with  $n > n^\#$  can be stable. This last remark may also help to solve the following puzzle.

Secondly, as Rutz (2002) points out, in almost all IEAs that he investigated the number of participants exceeds the minimum requirement – a phenomenon which cannot be explained by the models mentioned above. With my small modification, this observation could be accommodated easily.

Thirdly, given the assumptions on the payoff functions in these papers, I think it can easily be shown that the equilibrium participation clause is always full participation.<sup>47</sup> This is certainly not in line with real world observations and is probably caused by the assumption of symmetric players. Hence, it would be interesting to analyze a minimum participation clause in a richer model with heterogeneous countries.

---

<sup>47</sup> Carraro et al. (2004) derive a condition which, if violated, means that full participation is not an equilibrium choice of all countries. However, this condition will never be violated for their assumption on payoff functions as payoffs of signatories increase in the size of the coalition (noticed by the authors) but also those of non-signatories (not noticed by the authors). Thus, ex-post, the grand coalition is always the equilibrium choice of countries when proposing a minimum participation clause.

## References

- Altamirano-Cabrera, J.-C. (2007), On the Political Economy of International Climate Agreements. Ph.D. thesis, University of Wageningen, The Netherlands.
- Altamirano-Cabrera, J.-C. and M. Finus (2006), Permit Trading and Stability of International Climate Agreements. "Journal of Applied Economics", vol. 9, pp. 19-48.
- Altamirano-Cabrera, J.C., M. Finus, R. Dellink (2006), Do Abatement Quotas Lead to More Successful Climate Coalitions? Working Paper in Economics No. 383, University of Hagen, Germany, 2006. Forthcoming: "The Manchester School".
- Asheim, G.B., C.B. Froyen, J. Hovi and F.C. Menz (2006), Regional versus Global Cooperation for Climate Control. "Journal of Environmental Economics and Management", vol. 51(1), pp. 93-109.
- d'Aspremont, C., A. Jacquemin, J.J. Gabszewicz and J.A. Weymark (1983), On the Stability of Collusive Price Leadership. "Canadian Journal of Economics", vol. 16, pp. 17-25.
- Babiker, M.H. (2001). The CO<sub>2</sub> Abatement Game: Costs, Incentives, and the Enforceability of a Sub-global Coalition. "Journal of Economics & Control", vol. 25, pp. 1-34.
- Babiker, M.H., J.M. Reilly, M. Mayer, R.S. Eckaus, I. Sue Wing and R.C. Hyman (2001), The MIT Emissions Prediction and Policy Analysis (EPPA) Model: Revisions, Sensitivities and Comparison of Results', Joint program on the science and policy of global change Report 71, MIT, Cambridge.
- Barrett, S. (1994), Self-Enforcing International Environmental Agreements. "Oxford Economic Papers", vol. 46, pp. 878-894.
- Barrett, S. (1997), The Strategy of Trade Sanctions in International Environmental Agreements. "Resource and Energy Economics", vol. 19, pp. 345-361.
- Barrett, S. (1999), A Theory of Full International Cooperation. "Journal of Theoretical Politics", vol. 11(4), pp. 519-541.
- Barrett, S. (2002), Consensus Treaties. "Journal of Institutional and Theoretical Economics", vol. 158, pp. 529-547.
- Barrett, S. (2003), Environment and Statecraft: The Strategy of Environmental Treaty-making. Oxford University Press, New York.
- Bloch, F. (1995), Endogenous Structures of Associations in Oligopolies. "RAND Journal of Economics", vol. 26, pp. 537-556.
- Bloch, F. (2003), Non-cooperative Models of Coalition Formation in Games with Spillovers. In: Carraro, C. (ed.), The Endogenous Formation of Economic Coalitions. Edward Elgar, Cheltenham, UK et al., ch. 2, pp. 35-79.
- Böhringer, C. and A. Löschel (2006), Computable General Equilibrium Models for Sustainability Impact Assessment : Status Quo and Prospects, vol. 60, pp. 49-64.
- Böhringer, C. and C. Vogt (2004), The Dismantling of a Breakthrough: The Kyoto Protocol as Symbolic Policy. "European Journal of Political Economy", vol. 20, pp. 597-618.
- Bosello, F., B. Buchner and C. Carraro (2003), Equity, Development, and Climate Change Control. "Journal of the European Economic Association", vol. 1(2-3), pp. 601-611.

- Buchner, B. and C. Carraro (2005a), Modelling Climate Policy, Perspectives on Future Negotiations. "Journal of Policy Modeling", vol. 27, pp. 711-732.
- Buchner, B. and C. Carraro (2005b), Regional and Sub-global Climate Blocs. A Game-theoretic Perspective on Bottom-up Climate Regimes. Working Paper No. 21.2005, Fondazione Eni Enrico Mattei, Milan, Italy.
- Bucholz, W., A. Haupt, W. Peters (2005), International Environmental Agreements and Strategic Voting. "Scandinavian Journal of Economics", vol. 107, pp. 175-195.
- Caparrós, A., J.-C. Péreau and T. Tazdaït (2004), North-South Climate Change Negotiations: A Sequential Game with Asymmetric Information. "Public Choice", vol. 121, pp. 455-480.
- Carraro, C. (2000), Roads towards International Environmental Agreements. In: Siebert, H. (ed.), The Economics of International Environmental Problems. Mohr Siebeck, Tübingen, pp. 169-202.
- Carraro, C., J. Eyckmans, M. Finus (2006), Optimal Transfers and Participation Decisions in International Environmental Agreements. "Review of International Organizations", vol. 1, pp. 379-396.
- Carraro, C. and C. Marchiori (2003), Stable Coalitions. In: Carraro, C. (ed.), The Endogenous Formation of Economic Coalitions. Edward Elgar, Cheltenham, UK et al., ch. 5, pp. 156-198.
- Carraro, C., C. Marchiori and S. Oreffice (2004), Endogenous Minimum Participation in International Environmental Treaties. CEPR Discussion Paper No. 4281, London.
- Carraro, C. and D. Siniscalco (1993), Strategies for the International Protection of the Environment. "Journal of Public Economics", vol. 52(3), pp. 309-328.
- Carraro, C. and D. Siniscalco (1997), R&D Cooperation and the Stability of International Environmental Agreements. In: Carraro, C. (ed.), International Environmental Negotiations: Strategic Policy Issues. Edward Elgar, Cheltenham, UK et al., ch. 5, pp. 71-96.
- Carraro, C. and D. Siniscalco (1998), International Environmental Agreements: Incentives and Political Economy. "European Economic Review", vol. 42(3-5), pp. 561-572.
- Chander, P. and H. Tulkens (1992), Theoretical Foundations of Negotiations and Cost Sharing in Transfrontier Pollution Problems. "European Economic Review", vol. 36(2-3), pp. 388-399.
- 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(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(3), pp. 379-401.
- Dellink, R.B., M. Finus, E.C. van Ierland and J.-C. Altamirano (2004), Empirical background paper of the STACO model, available on the STACO website <http://www.enr.wur.nl/uk/staco>, Wageningen University.
- Dellink, R., M. Finus, N. Oliemann (2005), Coalition Formation under Uncertainty: The Stability Likelihood of an International Climate Agreement. Working Paper No. 98.2005,

- Fondazione Eni Enrico Mattei, Italy. Forthcoming: "Journal of Environmental and Resource Economics".
- Diamantoudi E., Sartzetakis E. (2006), Stable International Environmental Agreements: an Analytical Approach. "Journal of Public Economic Theory", vol. 8; pp. 247-263.
- Ecchia, G. and M. Mariotti (1998), Coalition Formation in International Environmental Agreements and the Role of Institutions. "European Economic Review", vol. 42(3), pp. 573-582.
- Endres, A. (1997), Negotiating a Climate Convention – The Role of Prices and Quantities. "International Review of Law and Economics", vol. 17(1), pp. 147-156.
- Endres, A. and C. Ohl (2000), Taxes versus Quotas to Limit Global Environmental Risks: New Insights into an Old Affair. "Environmental Economics and Policy Studies", vol. 3, pp. 399-423.
- Endres, A. and C. Ohl (2003), International Environmental Cooperation with Risk Aversion. "International Journal of Sustainable Development", vol. 6, pp. 378-392.
- Eyckmans, J. (1997), Nash Implementation of a Proportional Solution to International Pollution Control Problems. "Journal of Environmental Economics and Management", vol. 33(3), pp. 314-330.
- Eyckmans, J. and M. Finus (2006a), New Roads to International Environmental Agreements: The Case of Global Warming. "Environmental Economics and Policy Studies", vol. 7, pp. 391-414.
- Eyckmans, J. and M. Finus (2006b), Coalition Formation in a Global Warming Game: How the Design of Protocols Affects the Success of Environmental Treaty-Making. "Natural Resource Modeling", vol. 19, pp. 323-358.
- Eyckmans, J. and M. Finus (2007), Measures to Enhance the Success of Global Climate Treaties. "International Environmental Agreements", vol. 7, pp. 73-97.
- Eyckmans, J. and H. Tulkens (2003), Simulating Coalitionally Stable Burden Sharing Agreements for the Climate Change Problem. "Resource and Energy Economics", vol. 25(4), pp. 299-327.
- Farrell, J. and E. Maskin (1989), Renegotiation in Repeated Games. "Games and Economic Behavior", vol. 1(4), pp. 327-360.
- Finus, M. (2001), Game Theory and International Environmental Cooperation, Edward Elgar, Cheltenham, UK.
- Finus, M. (2003a), Stability and Design of International Environmental Agreements: The Case of Transboundary Pollution. In: Folmer, H. and T. Tietenberg (eds.), International Yearbook of Environmental and Resource Economics, 2003/4, Edward Elgar, Cheltenham, UK, ch. 3, pp. 82-158.
- Finus, M. (2003b), New Developments in Coalition Theory: An Application to the Case of Global Pollution. In: Marsiliani, L., M. Rauscher and C. Withagen (eds.), Environmental Policy in an International Perspective, Kluwer, Dordrecht, Holland, pp. 19-49.
- Finus, M. (2004), Modesty Pays: Sometimes! Working Paper No. 68.04, Fondazione Eni Enrico Mattei, Milano.

- Finus, M. (2007), Possibilities for Cooperation in International Pollution Control. In: Cognoy, M. and K.W. Steininger (eds.), *The Economics of Global Environmental Change International Cooperation for Sustainability*, Edward Elgar, Cheltenham U.K., ch. 6, pp. 157-195.
- 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.
- Finus, M. and B. Rundshagen (2003), Endogenous Coalition Formation in Global Pollution Control: A Partition Function Approach. In: Carraro, C. (ed.), *Endogenous Formation of Economic Coalitions*, Edward Elgar, Cheltenham, UK, ch. 6, pp. 199-243.
- Finus, M. and B. Rundshagen (2006a), A Micro-Foundation of Core-Stability in Positive Externality Coalition Games. „*Journal of Institutional and Theoretical Economics*“, vol. 162, pp. 329-346.
- Finus, M. and B. Rundshagen (2006b), Participation in International Environmental Agreements: The Role of Timing and Regulation. “*Natural Resource Modeling*”, vol. 19, pp. 165-200.
- Finus, M., E. van Ierland and J.C. Altamirano-Cabrera (2005), The Effect of Membership Rules and Voting Schemes on the Success of International Climate Agreements. “*Public Choice*”, vol. 125, pp. 95-127.
- Finus, M., E. van Ierland and R. Dellink (2006), Stability of Climate Coalitions in a Cartel Formation Game. “*Economics of Governance*”, vol. 7, pp. 271-291.
- Finus, M., M.E. Sáiz and E.M.T. Hendrix (2004), An Empirical Test of New Developments in Coalition Theory for the Design of International Environmental Agreements. Mansholt Working Paper No. 14, University of Wageningen, The Netherlands.
- Finus, M. and S. Tjøtta (2003), The Oslo Protocol on Sulfur Reduction: The Great Leap Forward? *Journal of Public Economics*, vol. 87, pp. 2031-2048.
- Folmer, H. and P. van Mouche (2000), Transboundary Pollution and International Cooperation. In: Folmer, H. and T. Tietenberg (eds.), *The International Yearbook of Environmental and Resource Economics*, UK et al., Edward Elgar, Cheltenham, pp. 231-267.
- Folmer, H. and P. van Mouche (1994), Interconnected Games and International Environmental Problems II. "Annals of Operations Research", vol. 54, pp. 97-117.
- Folmer, H., P. van Mouche and S.E. Ragland (1993), Interconnected Games and International Environmental Problems. "Environmental and Resource Economics", vol. 3(4), pp. 313-335.
- Folmer, H. and A. de Zeeuw (1999), Game Theory in Environmental Policy Analysis. In: van den Bergh, J.C. (ed.), *Handbook of Environmental and Resource Economics*. Edward Elgar, Cheltenham, UK et al., ch. 70, pp. 1089-1098.
- Folmer, H. and A. de Zeeuw (2000), International Environmental Problems and Policy. In: Folmer, H. and H.L. Gabel (eds.), *Principles of Environmental and Resource Economics: A Guide for Students and Decision-Makers*, UK et al., Edward Elgar, Cheltenham, pp. 447-478.

- Fujita, T. (2004), Design of International Environmental Agreements und Uncertainty. "Environmental Economics and Policy Studies", vol. 6, pp. 103-118.
- Golombek, R. and M. Hoel (2005), Climate Policy under Technology Spillovers. "Environmental and Resource Economics", vol. 31, pp. 201-227.
- Gua, J., Hepburn, C. J., R. S.J. Tol and D. Anthoff (2006), Discounting and the Social Cost of Carbon: A Closer Look at Uncertainty. "Environmental Science & Policy", vol. 9, pp. 205-216.
- Haeringer G. and Courtois, P. (2005), The Making of International Environmental Agreements. CODE Working Paper 652.05.
- Hoel, M. (1992), International Environment Conventions: The Case of Uniform Reductions of Emissions. "Environmental and Resource Economics", vol. 2(2), pp. 141-159.
- Hoel, M. and K. Schneider (1997), Incentives to Participate in an International Environmental Agreement. "Environmental and Resource Economics", vol. 9(2), pp. 153-170.
- Jeppesen, T. and P. Andersen (1998), Commitment and Fairness in Environmental Games. In: Hanley, N. and H. Folmer (eds.), Game Theory and the Environment. Edward Elgar, Cheltenham, UK et al, ch. 4, pp. 65-83.
- Kempfert, C. (2005), Induced Technological Change in a Multi-regional, Multi-sectoral, Integrated Assessment Model (WIAGEM): Impact Assessment of Climate Policy Strategies. Special Issue of "Ecological Economics", vol. 54, pp. 293-305.
- Kolstad, C.D. (1996), Fundamental Irreversibilities in Stock Externalities. "Journal of Public Economics", vol. 60, pp. 221-233.
- Kolstad, C.D. (2007), Systematic Uncertainty in Self-enforcing International Environmental Agreements. "Journal of Environmental Economics and Management", vol. 53, pp. 68-79.
- Kolstad, C.D. and A. Ulph (2006), International Environmental Agreements under Uncertainty when Learning Reveals Differences between Countries. Paper presented at the conference "Learning and climate change", IIASA, Laxenburg, Austria, 10-11 April 2006.
- Lange, A. (2006), The Impact of Equity-preferences on the Stability of International Environmental Agreements. "Environmental and Resource Economics", vol. 34(2), pp. 247-267.
- Lange, A. and C. Vogt (2003), Cooperation in International Environmental Negotiations due to a Preference for Equity. "Journal of Public Economics", vol. 87(9-10), pp. 2049-2067.
- Lange, A., C. Vogt and A. Ziegler (2007), On the Importance of Equity in International Climate Policy: An Empirical Analysis. "Energy Economics", vol. 29, pp. 545-562.
- Michaelowa, A. (1998), Climate Policy and Interest Groups – a Public Choice Analysis. "Intereconomics", vol. 33(6), pp. 251-259.
- Murdoch, J.C. and T. Sandler (1997a), Voluntary Cutbacks and Pretreaty Behavior: The Helsinki Protocol and Sulfur Emissions. "Public Finance Review", vol. 25(2), pp. 139-162.
- Murdoch, J.C. and T. Sandler (1997b), The Voluntary Provision of a Pure Public Good: The Case of Reduced CFC Emissions and the Montreal Protocol. "Journal of Public Economics", vol. 63(3), pp. 331-349.

- Na, S.-L. and H.S. Shin (1998), International Environmental Agreements under Uncertainty. "Oxford Economic Papers", vol. 50, pp. 173-185.
- Nordhaus, W.D. and Z. Yang (1996), A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies. "American Economic Review", vol. 86(4), pp. 741-765.
- Peters W. Schuler, C. (2006), International Environmental Agreements: Can Equity and Fairness Enlarge the Size of the Coalition? Paper presented at the "6<sup>th</sup> Meeting on Game Theory and Practice", Zaragoza, Spain, July 10<sup>th</sup>-12<sup>th</sup>, 2006.
- Ray, D. and R. Vohra (2001), Coalitional Power and Public Goods. "Journal of Political Economy", vol. 109, pp. 1355-1384.
- Rose, A., B. Stevens, J. Edmonds and M. Wise (1998), International Equity and Differentiation in Global Warming Policy: An Application to Tradeable Emission Permits. "Environmental and Resource Economics", vol. 12(1), pp. 25-51.
- Rubio, S. J. and B. Casino (2005), Self-enforcing International Environmental Agreements with a Stock Pollutant. "Spanish Economic Review", vol. 7(2), pp. 89-109.
- Rubio S. J., Ulph A. (2002) A Simple Dynamic Model of International Environmental Agreements with a Stock Pollutant. Discussion Papers in Economics, University of Southampton, No. 209.
- Rubio S. J., Ulph A. (2006), Self-enforcing International Environmental Agreements Revisited. "Oxford Economic Papers", vol. 58, pp. 233-263.
- Rubio S. J., Ulph A. (2007), An Infinite-horizon Model of Dynamic Membership of International Environmental Agreements. Forthcoming: Journal of Environmental Economics and Management.
- Rutz, S. (2002), International Environmental Agreements: Much Ado about Nothing? Ph.D. thesis, Shaker Verlag, Aachen.
- Scriciu, S.S. (2006), The Inherent Dangers of Using Computable General Equilibrium Models as a Single Integrated Modelling Framework for Sustainability Impact Assessment. A Critical Note on Böhringer and Löschel (2006), vol. 4, pp. 678-684.
- Swanson, T. and R. Mason (2003), The Impact of International Environmental Agreements: The Case of the Montreal Protocol. In: Marsiliani, L., M. Rauscher and C. Withagen (eds.), Environmental Policy in an International Perspective. Kluwer Academic Publishers, Dordrecht et al., pp. 51-80.
- Tol, R.S.J. (2001), Climate Coalitions in an Integrated Assessment Model. "Computational Economics", vol. 18(2), pp. 159-172.
- Tulkens, H. (1998), Cooperation versus Free-Riding in International Environmental Affairs: Two Approaches. In: Hanley, N. and H. Folmer (eds.), Game Theory and the Environment. Edward Elgar, Cheltenham, UK et al., ch. 2, pp. 30-44.
- Ulph, A. (1998), Learning about Global Warming? In: Hanley, N. and H. Folmer (eds.), Game Theory and the Environment. Edward Elgar, Cheltenham, UK et al., ch. 13, pp. 255-286.

- Ulph, A. (2004), Stable International Environmental Agreements with a Stock Pollutant, Uncertainty and Learning. "Journal of Risk and Uncertainty", vol. 29(1), pp. 53-73.
- Ulph, A. and D. Maddison (1997), Uncertainty, Learning and International Environmental Policy Coordination. "Environmental and Resource Economics", vol. 9, pp. 451-466.
- Ulph, A. and D. Ulph (1997), Global Warming, Irreversibility and Learning. "The Economic Journal", vol. 107, pp. 636-650.
- Wagner, U.J. (2001), The Design of Stable International Environmental Agreements: Economic Theory and Political Economy. "Journal of Economic Surveys", vol. 15 (3), pp. 377–411.
- Weikard, H.-P., Finus, M. and J.C. Altamirano-Cabrera (2006), The Impact of Surplus Sharing on the Stability of International Climate Coalitions. "Oxford Economic Papers", vol. 58, pp. 209-232.
- De Zeeuw, A. (2005), Dynamic Effects on the Stability of International Environmental Agreements. Working Paper No. 41.2005, Fondazione Eni Enrico Mattei.

**Table 1: Structure of Coalition Formation in Membership Models\***

**1. Stage: participation**

sequence	<u>simultaneous</u>	sequential no revision of members      revision of members
agreements	<u>single</u>	<i>multiple</i>
membership	<u>open</u>	<i>exclusive</i> <i>majority</i> <i>unanimity</i>

**2. Stage: abatement and transfers**

sequence	<u>simultaneous (Cournot)</u>	sequential (Stackelberg)
abatement	<u>joint welfare maximization</u> <u>(efficient)</u>	<i>bargaining</i> <i>majority</i> <i>unanimity</i>
transfers	<u>no</u>	<i>yes</i>
payoffs	<u>objective</u> <u>social planner</u> <u>material</u> <u>certain</u>	subjective politician non-material uncertain

\* Base assumptions are underlined, alternative assumptions as reported in Table 2, 3, 4 and 5 in section 3 are indicated italic. All other alternative assumptions are discussed in section 4.

**Table 2: Membership Rules and Multiple Coalitions: Stable Coalition Structures without Transfers\***

Coalition Structure	OM		EM-MV		EM-UV		Global Abatement % BAU	Global Welfare bln US\$
	S	M	S	M	S	M		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
grand coalition (global optimum)	n	n	n	n	n	n	21.4	6,031
industrialized countries	n	n	n	n	n	n	8.9	3,140
industrialized countries without USA	n	n	n	n	n	n	6.9	2,692
{O-OECD,India,Brazil},{FSU,ROW},{EEX,China}	-	n	-	n	-	y	5.8	2,452
{FSU,Brazil,ROW},{O-OECD,India},{EEX,China}	-	n	-	n	-	y	5.8	2,451
{O-OECD,India,Brazil},{FSU,ROW},{China,DAE}	-	n	-	n	-	y	5.7	2,415
{FSU,Brazil,ROW},{O-OECD,India},{China,DAE}	-	n	-	n	-	y	5.7	2,414
{O-OECD,India,Brazil},{FSU,ROW},{EEX,DAE}	-	n	-	n	-	y	5.4	2,263
{FSU,Brazil,ROW},{O-OECD,India},{EEX,DAE}	-	n	-	n	-	y	5.4	2,262
{FSU,Brazil,ROW},{O-OECD,DAE},{EEX,India}	-	n	-	n	-	y	5.4	2,261
{FSU,Brazil,ROW},{O-OECD,EEX},{India,DAE}	-	n	-	n	-	y	5.3	2,255
{EEX,China}	n	n	n	n	y	n	5.2	2,190
{China,DAE}	n	n	n	n	y	n	5.1	2,153
{FSU,Brazil,ROW}	n	n	y	n	y	n	5.0	2,131
singleton coalition structure (Nash equilibrium)	y	y	y	y	y	y	4.6	1,960

\* Base assumptions as indicated in Table 1 if not stated otherwise. OM = open membership, EM-MV = exclusive membership majority voting, EM-UV = exclusive membership unanimity voting, S = single agreement, M = multiple agreement, n = not stable, y = stable; Global Welfare = billion US dollars discounted with a discount rate of 2 % over the period 2010-2110. Global Abatement = average abatement in percentages from the Business as Usual Scenario over the period 2010-2110; Industrialized countries include: USA, Japan, EU-15, O-OECD, EE and FSU; abbreviations of regions as explained in the text in section 3.1.

**Table 3: The Role of Sharing Scheme: Stable Coalitions with Transfers\***

Transfer Scheme	Coalition Members	Global Abatement		Global Welfare
		% BAU		bln US\$
(1)	(2)	(3)		(4)
Global Optimum	grand coalition <sup>#</sup>	21.4		6,031
1) GDP	EU-15, China	7.3		2,942
2) Emissions	USA, EE, EEX, China	8.6		3,418
3) Equal Sharing	EE, China, India	5.9		2,499
4) Ability to Pay	China, India	5.5		2,334
5) Inverse Emissions	EE, Brazil	4.6		1,981
	China, Brazil	4.8		2,059
6) Population	EEX, China	5.2		2,190
Nash Equilibrium	singleton coalition structure	4.6		1,960

\* Base assumptions as indicated in Table 1 if not stated otherwise. For the general construction of outcome-based transfer schemes, see footnote 18. The weights for the six transfer schemes are defined as follows: GDP:  $\lambda_i = \text{GDP}_i / \sum_{j \in S} \text{GDP}_j$ ; Emissions:  $\lambda_i = E_i / \sum_{j \in S} E_j$ ; Equal Sharing:  $\lambda_i = 1/\#S$ ; Ability to Pay:  $(\text{GDP}_i / \text{POP}_i)^{-1} / \sum_{j \in S} (\text{GDP}_j / \text{POP}_j)^{-1}$ ; Inverse Emissions:  $\lambda_i = E_i^{-1} / \sum_{j \in S} (E_j)^{-1}$ ; Population:  $\lambda_i = \text{POP}_i / \sum_{j \in S} \text{POP}_j$  with GDP=gross domestic product, E=BAU-emissions, #S= size of coalition S, POP=population; all values refer to 2010, see Weikard et al. (2006) for details. Abbreviations of regions as explained in the text in section 3.1. # = no stable coalition. Global Abatement and Global Welfare as defined in Table 2.

**Table 4: The Role of Emission Reduction Quotas and Bargaining: Stable Coalitions without Trading\***

Agreement Design	Coalition Members	Global Abatement	Global Welfare
		% of BAU	bln US\$
Global Optimum	Grand coalition <sup>#</sup>	21.4	6,031
1) Efficient Design	---	4.6	1,960
2) Common Quota	India, ROW	5.0	2,100
3) Median Quota Proposal	India, ROW	5.0	2,103
4) Lowest Quota Proposal	India, ROW	4.8	2,088
	EU-15, China, India, ROW	6.8	2,725
			} (2,407)
Nash Equilibrium	singleton coalition structure	4.6	1,960

\* Agreement Design 1) to 4) as defined in the text in section 3.2. # = no stable coalition. Global Abatement and Global Welfare as defined in Table 2.

**Table 5: The Role of Emission Quotas and Bargaining: Stable Coalitions with Trading\***

Agreement Design	Coalition Members	Global Abatement % of BAU	Global Welfare bln US\$
Global Optimum	grand coalition <sup>#</sup>	21.4	6,031
1) Efficient Design	---	4.6	1,960
2) Common Quota	India, ROW <sup>I</sup>	5.0	2,104
	Japan, EU-15, China, India	7.3	2,982
3) Median Quota Proposal	India, ROW <sup>I</sup>	5.0	2,104
	Japan, India, ROW	5.5	2,261
	Japan, EU-15, China, India	7.3	2,973
	USA, Japan, China, India	8.3	3,307
	USA, EU-15, China, India	9.0	3,492
4) Lowest Quota Proposal	India, ROW <sup>I</sup>	4.8	2,088
	EU-15, China, India, ROW <sup>I</sup>	6.8	2,811
	Japan, EU-15, China, India, ROW	7.5	3,050
	USA, Japan, EU-15, China, India	8.4	3,350
Nash Equilibrium	singleton coalition structure	4.6	1,960

\* Agreement Design 1) to 4) as defined in the text in section 2.2. # = no stable coalition; I = only internally stable coalition. Global Abatement and Global Welfare as defined in Table 2.