Strategic Behavior and the Environment, 2014, 4: 361–392

The Role of Emissions Trading and Permit Allocation in International Climate Agreements with Asymmetric Countries

Michael Jakob,[1,2] Kai Lessmann,[2] and Theresa Wildgrube[3]

[1]*Mercator Research Institute on Global Commons and Climate Change (MCC), Berlin, Germany; jakob@mcc-berlin.net*
[2]*Potsdam Institute for Climate Impact Research (PIK), P.O. Box 601203, 14412 Potsdam, Germany*
[3]*Fundación IDEA, Calle Leibnitz # 20 Piso 7, Colonia Anzures Delegación Miguel Hidalgo, 11590 México*

ABSTRACT

This paper presents a model of international environmental agreements in which cooperation between asymmetric countries can arise through pure self-interest. It demonstrates how emissions trading creates economic surplus by exploiting asymmetries. This surplus can be distributed via the appropriate allocation of reduction commitments, which ensures that membership in the agreement is compatible with countries' incentives to join. While this mechanism improves upon the business-as-usual outcome, it does not solve the underlying collective action problem wherein abatement falls short of the social optimum. We also show that countries' incentives to participate in a global climate agreement crucially depend on the permit allocation schemes, and that allocation schemes that ensure full participation in the global climate agreement might be at odds with fundamental equity considerations.

ISSN 1944-012X; DOI 10.1561/102.00000050

Keywords: International environmental agreements; coalition game; emissions trading; allocation scheme.

JEL Codes: C72; Q54; H41

Introduction

As a common result, the early models of coalition stability that draw on non-cooperative game theory find that a global climate agreement can only sustain a low number of participants or low gains from cooperation, and identify strong free-riding incentives as the underlying reason (Hoel, 1992; Carraro and Siniscalco, 1993; Barrett, 1994). The setting of these studies has become a standard model for a broad literature on international environmental agreements (surveyed in Finus, 2008). This standard model is based on two fundamental assumptions: (i) the costs of providing a global public good and the associated benefits are the same for all players, and (ii) the coalition of players that implements an environmental agreement maximizes the joint welfare of all of its members. The latter implies that the optimal level of the public good is set collectively and the necessary contributions are allocated efficiently among the members. Yet, recent studies that relax either of these assumptions draw more optimistic conclusions regarding the feasibility of a global climate treaty (cf. discussion of the literature in the second section).

This paper presents the first analytically tractable model that simultaneously considers asymmetric countries and departs from the assumption of joint-welfare optimization. We adopt the standard framework where countries decide whether or not to join a coalition that provides a global public good based on individual benefits (i.e., internal and external stability of coalitions; cf. Carraro and Siniscalco, 1993). The countries are distinguished by type, which reflects their costs of contributing to the public good and the benefits derived from it. Rather than using joint-welfare maximization, their contributions depend on their country type and are determined by a rule that is not subject to negotiation. Furthermore, we assume that members of the coalition engage in the trade of emissions permits. This set-up is particularly relevant for the case of a global climate agreement in which one type of (industrialized) country (which can be expected to display a higher willingness to pay for climate change mitigation) finances abatement in another type of (developing) country.

Our results indicate that such asymmetries between countries can be exploited through emissions trading. When emissions trading is restricted to coalition members, the gains of trade (efficiency) add to the incentive for countries to join an international climate agreement. We show that even though this does not resolve the underlying collective action problem, it does mitigate it. The resulting agreement achieves higher abatement and payoffs than the business-as-usual case. Furthermore, stable coalitions in this framework can easily be large in number, contrary to more pessimistic results in the literature.

This paper proceeds as follows: The second section reviews the relevant literature. The third section presents the basic model of emissions trading with asymmetric countries. The fourth section investigates the equilibrium conditions of the model and the complementary roles taken by different types of countries. The fifth section compares the resulting stable coalitions to the non-cooperative equilibrium and the first-best outcome and contrasts our approach (which is based on an exogenous rule of contributions) with the standard assumption of joint-welfare maximization. It also elaborates on the role of the allocation of abatement and demonstrates how equity considerations regarding the initial distribution of commitments have the potential to impede the formation of a coalition. The final section discusses the policy implications of our results and provides concluding remarks.

Links to Previous Literature

The literature studying the formation of international environmental agreements (IEAs) has generated a rather pessimistic outlook on the provision of global public goods, most prominently in the context of mitigating climate change. The underlying rationale is that when a coalition fully internalizes all externalities, a larger coalition size implies higher contributions to the abatement jointly undertaken by the coalition for each individual country. This makes it less attractive for countries to join the coalition. Thus, only small coalitions are stable, and the resulting levels of the public good fall short of the socially optimal level (Carraro and Siniscalco, 1993; Barrett, 1994).[1] This line of research has traditionally adopted two major assumptions: (i) symmetry of players, i.e., all countries are assumed to be identical; and (ii) the coalition is assumed to maximize its members' joint welfare, i.e.,

[1] See Finus (2003, 2008) for a review of this large body of literature.

the sum of their net benefits. These assumptions have only been relaxed by a number of recent contributions that are summarized in Table 1 and are discussed in the following.

To relax the assumption of symmetric countries, asymmetries have been introduced in settings with and without side payments. Side payments allow coalition members to finance mitigation in countries that feature a lower willingness to pay, thus potentially stabilizing an agreement. In the absence of side payments, asymmetry has no substantial effect on cooperation (Barrett, 1997; Fuentes-Albero and Rubio, 2010; Colmer, 2011). However, side payments implemented via an appropriate scheme to divide costs and benefits among coalition members can significantly raise participation and bring the level of climate protection closer to the global optimum. Asymmetries then create opportunities for countries with a higher willingness to pay for the abatement of emissions to compensate countries with low mitigation costs, which consequently increases the incentive for both types to cooperate. In general, transfer schemes can take two forms: (i) a system of specific ex-ante abatement obligations in which cost-effectiveness is reinstalled by an emissions trading scheme (Altamirano-Cabrera and Finus, 2006; McGinty, 2007; Weikard *et al.*, 2006); (ii) a transfer scheme that ex-post distributes the economic surplus so that countries bearing a high burden are compensated for their efforts (Weikard, 2009; Carraro *et al.*, 2006; Botteon and Carraro, 1997, Biancardi and Villani, 2010).[2] Furthermore, Altamirano-Cabrera *et al.* (2008) have adopted a sophisticated approach that considers different voting rules, such as majority voting or unanimity, to decide on uniform abatement quotas that apply to all coalition members, thus showing that such procedures can, to a certain extent, dampen free-rider incentives.

When the assumption of joint-welfare maximization for the coalition is (at least partly) abandoned, cooperation will improve in the symmetric player setting and increase the levels of abatement. If the coalition adopts a less ambitious target than the one dictated by joint-welfare maximization, the incentive to free ride, as well as the costs of membership, may be reduced. In this case, larger coalitions that achieve more stringent levels of abatement become feasible. Barrett (2002) adopts the assumption of ''collective rationality,'' under which coalition members form so-called ''consensus treaties'' that maximize joint welfare *subject to participation by all countries.* If the

[2] Transfers have also been discussed for symmetric countries but in that setting, such transfers do little to increase coalition size and abatement (Carraro and Siniscalco, 1993).

Table 1. Overview of the literature, classified along the dimensions symmetric vs. asymmetric countries, and joint-welfare maximization vs. non-joint-welfare maximization.

	Joint-welfare Maximization	Non-joint-welfare Maximization
Symmetric countries	Standard assumption for analytically solved models (e.g., Hoel, 1992; Carraro and Siniscalco, 1993; Barrett, 1994).	Barrett (2002) for weakly collectively rational treaties and Finus and Maus (2008) for ''modesty'' constraints.
Asymmetric countries	**Analytical treatment:** Barrett (1997) for two types of asymmetry, McGinty (2007) for mean-preserving asymmetry, Weikard (2009) and Fuentes-Albero and Rubio (2010) for two types of asymmetry and Colmer (2011) for mean-preserving asymmetry. **Numerical analysis:** Botteon and Carraro (1997) for six regions, Altamirano-Cabrera and Finus (2006) for 12 regions, Carraro *et al.* (2006) for 6 regions, Weikard *et al.* (2006) for 12 regions, Altamirano-Cabrera *et al.* (2008) for 6 regions, Biancardi and Villani (2010) for two types of asymmetry.	**Analytical treatment:** Helm (2003) for endogenous permit choice. **Numerical analysis:** Carbone *et al.* (2009).

coalition anticipates the associated potential welfare gain due to broader participation, deviating from the optimal abatement level is indeed collectively rational. The result will then be a ''broad but shallow'' agreement that features broad participation but only relatively low levels of abatement for each individual coalition member. This outcome constitutes a contrast to the ''narrow but deep'' treaties predicted by earlier studies that rely on joint-welfare maximization. Finus and Maus (2008) follow a similar approach by assuming that the coalition aims at a less ambitious abatement target than the one that would be optimal for a given number of members. They find that in this constellation, free-rider incentives are reduced and larger coalitions are stable, which can — despite the lower amount of abatement undertaken by each individual coalition member — increase the overall abatement level (and hence, welfare).

These studies demonstrate that more optimistic results with regards to cooperation may occur when either the assumption of symmetric countries or the assumption of joint-welfare maximization is relaxed. To our knowledge, only very few studies have relaxed *both* of these assumptions simultaneously. Helm (2003) presents a model in which each country chooses its own reduction commitment, which is modeled as an endowment with emissions permits traded on an international carbon market. Countries that are more (or less) concerned with the environment will then choose lower (or higher) endowments of emissions permits if they are tradable, such that overall abatement may either increase or decline. Carbone *et al.* (2009) apply Helm's (2003) framework, in which each country chooses its own endowment with emissions permits, to a calibrated numerical model. Their estimates suggest that the trade of permits can incentivize cooperation between developed and developing countries.

Like Helm (2003) and Carbone *et al.* (2009), our paper studies asymmetric countries and non-joint-welfare maximization. However, it takes a different approach by analyzing cases in which these endowments are determined by rules that are not subject to negotiation. Therefore, it is closer, albeit not identical, to Finus and Maus (2008) where the coalition's level of ambition is determined in the pre-game stage. We argue that such rules can arise from, e.g., scientific findings or equity considerations, as will be explained in detail in the third section. In addition, while Carbone *et al.* (2009) derive their results from numerical calculations, our model is purely analytical. For this reason, this study illustrates how countries' incentives change as a general

function of benefits, abatement costs, and endowments with emission permits and allows for a detailed analysis of the involved economic mechanisms.

A Coalition Model of Emissions Trading Among Asymmetric Countries

This section first presents the actors' payoff functions and identifies the business-as-usual outcome, as well as the socially optimal levels of abatement. We then extend the base model by introducing emissions trading combined with an associated abatement commitment resulting from the allocation of emission permits that defines each country's responsibility for abating emissions.

Costs and Benefits

Let there be two types of countries — Northern (N-type) and Southern (S-type) — with linear benefits and quadratic abatement cost functions for the global public good, ''climate change mitigation,'' labeled e. Each country bears the costs of its own provision of e_i, but benefits from mitigation provided by all countries:

$$B_i = b_i \cdot \sum_{j \in \{N,S\}} e_j, \quad i = \{N, S\} \tag{1}$$

$$C_i = \frac{1}{2} c_i e_i^2, \quad i = \{N, S\} \tag{2}$$

This formulation of linear benefits and quadratic abatement costs is common in the literature (e.g., Weikard, 2009; Fuentes-Albero and Rubio, 2010). It constitutes a parsimonious representation of the idea that marginal costs will increase with higher levels of abatement because the least expensive mitigation options will be used up first. And while climate damages are commonly assumed to be a convex function of temperature change (e.g., quadratic in Nordhaus's 2008 DICE model), this is counteracted by temperature change being a concave function of emissions. Due to the interplay between the convexity of the former and the logarithmic shape of the latter function, a linear relationship can be regarded as at least a rough approximation of how climate damages (i.e., benefits from abatement) depend on emissions.[3]

[3] Own calculations, based on publicly available results from the DICE model (Nordhaus, 2008), reveal an almost linear relationship between cumulated emissions and cumulated discounted damages over the time horizon, 2005–2195.

The net benefits (i.e., ''welfare'') for each type of country {N, S} are simply derived from by difference between the benefits and costs:

$$W_i = B_i - C_i, \quad i = \{N, S\} \tag{3}$$

Furthermore, let there be N_N and N_S countries for each type, respectively.

The Business-As-Usual Outcome

Working autonomously (i.e., without a mechanism to establish cooperation between countries, which we call the ''business-as-usual case''), each country maximizes its individual net benefit by choosing its e_i such that its marginal costs equal its (private) marginal benefits:

$$b_i = c_i \cdot e_i^{BAU} \Rightarrow e_i^{BAU} = b_i/c_i, \quad i = \{N, S\}. \tag{4}$$

Hence, total abatement in the business-as-usual case is given by:

$$e_{tot}^{BAU} = N_N \cdot e_N^{BAU} + N_S \cdot e_S^{BAU} = N_N \cdot b_N/c_N + N_S \cdot b_S/c_S. \tag{5}$$

The Social Optimum

Summing up the net benefits of all countries yields the following expression for total welfare. Taking into account that with quadratic cost functions (i.e., increasing marginal abatement costs), all countries of one type will provide an identical amount of abatement in the social optimum, we have:[4]

$$\begin{aligned} W_{tot} = {} & N_N \cdot \left[b_N \cdot (N_N e_N + N_S e_S) - \frac{1}{2} c_N e_N^2 \right] \\ & + N_S \cdot \left[b_S \cdot (N_N e_N + N_S e_S) - \frac{1}{2} c_S e_S^2 \right]. \end{aligned} \tag{6}$$

From this, we can easily derive the socially optimal abatement efforts for both regions:

$$\begin{aligned} e_N^{opt} &= \frac{b_N N_N + b_S N_S}{c_N}, \quad \text{and} \\ e_S^{opt} &= \frac{b_N N_N + b_S N_S}{c_S}. \end{aligned} \tag{7}$$

[4] This expression implicitly assumes a utilitarian social welfare function, which is a standard assumption in the literature on coalition formation (cf. Barrett, 1994).

These expressions are quite straightforward: they simply state that in the social optimum, the marginal costs of abating one additional unit of emissions (i.e., $c_i \cdot e_i^{opt}$, $i = \{N, S\}$) equal the associated marginal social benefits that accrue to all countries (i.e., $b_N N_N + b_S N_S$).

Consequently, using (7), total abatement in the social optimum, which maximizes total welfare, can be expressed as:

$$e_{tot}^{opt} = N_N \cdot e_N^{opt} + N_S \cdot e_S^{opt} = (N_N/c_N + N_S/c_S) \cdot (b_N N_N / b_S N_S) \qquad (8)$$

Coalition with Emissions Trading

Let us now consider the case in which countries have the opportunity to enter into a global climate agreement with emissions trading, such that (i) marginal abatement costs across all members of the coalition[5] are equalized at permit price p and (ii) each N-type and S-type country that is a member of the coalition contributes a predetermined amount of emission abatement of o_N and o_S, respectively, which is achieved through a combination of domestic abatement and trading of emissions permits.

We assume that these reduction commitments are determined by a rule that lies outside of the scope of negotiations. Such rules could be derived from scientific findings or political targets that have been agreed upon in earlier negotiations. In addition, equity considerations, as well as the political feasibility of proposed rules, can be expected to play important roles. For instance, Altamirano-Cabrera and Finus (2006) examine various ''pragmatic'' and ''equitable'' schemes to allocate emissions permits. The former include allocations that are relative to historic or business-as-usual emissions, while the latter are either based on equal per-capita emissions rights, are inversely proportional to historical emissions, or are based on the ability to pay.[6] One manifestation of such a rule could be the widely endorsed goal to limit global warming to 2°C above the pre-industrial global mean temperature (cf. Jaeger and Jaeger, 2011) in combination with the IPCC's

[5] To keep the analysis tractable, we restrict this discussion to the case of a single coalition. See e.g., Asheim *et al.* (2006) for a recent discussion on a model featuring several (regional) climate agreements.

[6] Altamarino-Cabrera and Finus (2006) find that permit trading can raise participation and total abatement, with pragmatic schemes being more successful than equitable ones. As a possible extension to their analysis, they suggest dropping the assumption of joint-welfare maximization, as we have done in this paper.

(2007) recommendation that in order to reach this target, global emissions should decline by 50% in 2050, relative to the year 2000, with industrialized countries' emissions reduced by 80%–95%. In principle, these recommendations could be translated directly into benchmarks for contributions by each type of country. For instance, each developed country could be required to commit to reducing its emissions by, say, 90%, and each developing country by, say, 10%.

The underlying assumption is that countries can decide whether or not to join the agreement but have no opportunity to renegotiate its contributions once they are fixed in the agreement. To preclude renegotiation is in the vein of, for example, the concept of ''weakly collectively rational'' treaties, which exempt the level of punishment from joint-welfare maximization, i.e., renegotiation (Barrett, 2002). This constitutes a deliberate departure from the assumption of joint-welfare maximization that is common in the literature. It also differs from Carbone *et al.* (2009) who assume that each country chooses its own contribution. In our view, Carbone *et al.* (2009) and our study can be regarded as polar-opposite cases of complete freedom of choice of individual contributions and no influence at all. While Carbone *et al.* (2009) assume that countries are in no way constrained in their emissions targets and, thus, may overestimate their freedom of choice, they probably have more leeway to influence emissions targets than admitted in our approach.

Note that because any reduction commitment can be regarded conversely as an allowance of how much can be emitted, we use the terms ''reduction commitment'' and ''allocation of emissions permits'' interchangeably for the remainder of this paper. The number of countries of each type participating in the agreement is denoted by n_N and n_S, respectively. Only member countries are allowed to engage in permit trading. At permit price p, the domestic abatement level for each country participating in the agreement is determined by the condition that its marginal abatement costs ($c_i e_i^C$, $i = \{N, S\}$) equal the permit price. Hence, we can express the abatement undertaken by a member of the coalition (N-type or S-type) as a function of the permit price:

$$e_i^C = p/c_i, \quad i = \{N, S\}. \tag{9}$$

From this expression, we can derive the carbon price, which balances the supply of and demand for abatement by taking into account the fact that their reduction commitments require each N-type (or S-type) country to

abate o_N (o_S) units of emissions:

$$n_N \cdot p/c_N + n_S \cdot p/c_S = n_N \cdot o_N + n_S \cdot o_S, \tag{10}$$

which results in the following expression for the permit price p:

$$p = \frac{n_N o_N + n_S o_S}{n_N/c_N + n_S/c_S}. \tag{11}$$

We adopt the notation $x = n_N/n_S$ for the ratio of country types in the coalition. Then, when focusing on coalitions that include at least one country of each type (i.e., $n_N > 0$ and $n_S > 0$ such that $1/N_S \leq x \leq N_N$), the permit price can be rewritten as[7]

$$p(x) = \frac{x \cdot (o_N c_N) \cdot c_S + (o_S c_S) \cdot c_N}{x \cdot c_S + c_N}. \tag{11'}$$

Equation (11′) establishes the coalitional carbon price $p(x)$ as a function of the composition of the coalition.[8] Note that $x = n_N/n_S$ is discrete, i.e., $x \in \{\frac{n_N}{n_S}\colon\ n_N \in \{1, \ldots, N_N\}, n_S \in \{1, \ldots, N_S\}\}$. However, the function $p(x)$ is continuous and differentiable.

In order for our model to be relevant for the case of an international climate agreement, we adopt the following three assumptions.

A1 (benefit asymmetry): The benefits of N-type countries exceed those of S-type countries, i.e., $b_N > b_S$.

A2 (abatement above BAU): For both types of countries, abatement commitments under the climate agreement exceed abatement undertaken in the business-as-usual case, i.e., $o_i > b_i/c_i, i = \{N, S\}$.

A3 (cost asymmetry): N-type countries' marginal costs of meeting their reduction commitments by pure domestic mitigation are higher than that of S-type countries, i.e., $o_N c_N > o_S c_S$.

---

[7] We will see later that only coalitions including a non-zero number of N- and S-type countries are in accordance with assumptions $A1$—$A3$ specified below. Therefore, a coalition consisting of only N-type (or S-type) countries would result in a price of $p = o_N c_N$ ($p = o_S c_S$), which violates Observation 2.

[8] The permit price only depends on the ratio of the country types and not on the number of participating countries. As in our model, the amount of individual emissions reductions is already determined, i.e., the reduction commitment for any country does not increase with the number of countries that are members of the coalition (as it would under joint-welfare maximization). The permit price remains the same even with increased participation as long as the demand for permits by additional N-types is matched by permits supplied by additional S-types such that the ratio of N-types to S-types is the same.

These assumptions are quite straightforward. A1 simply ensures that there is heterogeneity between countries with regard to their benefits; hence, there is a ''benefit asymmetry'' assumption.[9] The ''abatement above BAU'' assumption (A2) excludes those cases in which participation in the international climate agreement is trivially fulfilled, as for $o_i < b_i/c_i$ where the required reduction would not go beyond the abatement that would be performed autonomously. Finally, according to the ''cost asymmetry'' assumption (A3), we only consider cases in which the marginal costs of meeting their reduction commitments by means of purely domestic abatement of emissions are higher for the country receiving higher benefits. Due to the differences in marginal abatement costs, opportunities to create economic surplus from emissions trading arise. As we will demonstrate in Observation 3, the cost asymmetry assumption ensures that reduction commitments are defined such that countries with a higher willingness to pay for climate change mitigation (i.e., higher benefits) buy emissions reductions from countries with lower benefits.

Using these assumptions, three observations can be made that will be useful for the further analysis of the coalition game.

Observation 1 *The price of emissions permits rises with the share of N-type countries and falls with the share of S-type countries in the coalition, i.e.,* $\frac{\Delta p}{\Delta x} > 0$.

According to the cost asymmetry assumption (A3), N-type countries display higher marginal abatement costs than S-type countries in terms of their respective reduction commitments. Intuitively, in a global carbon market, the equilibrium price of emissions permits has to settle somewhere between the highest and the lowest marginal abatement costs that would result if all abatement were performed domestically:

Observation 2 $o_S c_S < p < o_N c_N$*, i.e., the price of emissions permits has upper and lower limits, which are defined by the marginal abatement costs for the respective reduction commitments.*

[9] If N-types are industrialized countries, the assumption that they have a higher willingness to pay for avoiding climate damages can be justified by the fact that these countries dispose of more wealth that would be affected by climate impacts and may have more pronounced environmental awareness (e.g., Biancardi and Villani, 2010; Hannesson, 2010).

This gradient in costs determines the roles of N-type and S-type regions as buyers and sellers, respectively:

Observation 3 *N-type countries' actual abatement is below their reduction commitments, such that emissions trading results in transfer payments to S-type countries. Conversely, S-type countries' actual abatement exceeds their reduction commitments, such that they receive revenues from emissions trading.*

Observations 1, 2, and 3 follow from the model definition, as well as assumptions A1 to A3 (see Appendix for details).

Model Outcome: Equilibrium Coalition

Incentives to Join the Climate Agreement

This section discusses how countries' incentives to participate in an international climate agreement are determined by abatement costs, benefits, and reduction commitments. For the purpose of this paper, we regard an international climate agreement as a stable coalition of countries that meet their reduction commitments $\{o_N, o_S\}$ by any combination of domestic abatement and emissions trading. It should be noted that in our framework of reduction commitments that are established by non-negotiable rules, countries do not behave cooperatively in the sense of internalizing external effects on other coalition members (as is the case under joint-welfare maximization). Rather, coalition membership is driven by either the possibility to achieve emissions reductions at lower costs (N-types) or to receive revenues from the sale of emissions permits (S-types). Thus, the type of cooperation associated with coalition membership consists of participating in (mutually beneficial) emissions trading.

The incentives to join a coalition or not are summarized by the so-called stability function ϕ, which evaluates the net benefits of becoming a member of a coalition against the net benefits of remaining a non-member. If a country stays out of a coalition in which n_N and n_S countries (of each respective type) already participate, its welfare maximization problem results in abatement that is equal to the business-as-usual level specified in (4).[10] It enjoys

[10] This is due to the linear benefit function, which yields constant marginal benefits such that country i's marginal benefits from its own abatement efforts are independent from all other countries' abatement efforts.

the benefits of abatement of o_N and o_S by each of the n_N and n_S countries that are part of the coalition, respectively, and the $(N_N - n_N)$ and $(N_S - n_S)$, which continue to abate at business-as-usual levels. Hence, non-members' welfare W_i^{nm} is given by benefits minus mitigation costs:

$$W_i^{nm} = b_i \cdot [n_N o_N + n_S o_S + (N_N - n_N) \cdot e_N^{BAU} + (N_S - n_S) \cdot e_S^{BAU}] - \frac{1}{2} c_i (e_i^{BAU})^2, \quad i = \{N, S\}. \tag{12}$$

If a country joins the coalition, it (i) enjoys the additional benefits brought about by its own contribution to the coalition, (ii) incurs costs for domestic abatement $e_i^C = p/c_i$, $i = \{N, S\}$, and (iii) receives or provides transfer payments from emissions trading that are proportional to the difference between its reduction commitment and its domestic abatement (i.e., $p \cdot (e_i - o_i)$, $i = \{N, S\}$).

Therefore, the net benefits of being a member of a coalition with n_N and n_S members of N- and S-type, respectively, are:

$$W_i^{coal} = b_i \cdot [n_N o_N + n_S o_S + (N_N - n_N) \cdot e_N^{BAU} + (N_S - n_S) \cdot e_S^{BAU}] - \frac{1}{2} \cdot c_i \cdot \left(\frac{p}{c_i}\right)^2 + p \cdot \left(\frac{p}{c_i} - o_i\right), \quad i = \{N, S\}. \tag{13}$$

Thus, the stability function ϕ_i, which describes the incentives of being a member of the coalition compared to free riding, is given by:

$$\phi_i(n_i) = W_i^{coal}(n_i) - W_i^{nm}(n_i - 1), \quad i = \{N, S\}, \tag{14}$$

i.e., as the difference between the net benefits of each member of type i of the coalition with n_i members and the net benefit of each free-rider of type i with a coalition containing $n_i - 1$ countries of this type. Therefore, using (12) and (13) yields:

$$\phi_i(n_i) = b_i \cdot (o_i - b_i/c_i) + p \cdot (p/c_i - o_i) - \frac{1}{2} \cdot c_i \cdot (p/c_i)^2 + \frac{1}{2} \cdot c_i \cdot (b_i/c_i)^2, \quad i = \{N, S\}. \tag{14'}$$

Expression (14′) is straightforward: the first term describes the benefits of additional abatement compared to the business-as-usual case; the second term denotes the costs or revenues arising from emissions trading; and the

third and fourth terms stand for a coalition member's additional abatement costs, relative to the business-as-usual case. Note that the absolute number of countries of each type i (i.e., n_N and n_S) that are members of the coalition does not explicitly enter the stability function. However, the ratio of N-type and S-type countries determines the price of emissions permits, as shown in (11), and the price determines the incentive compatibility.

The following observation captures how countries' incentives to become members of the coalition depend on the price of emissions permits:

Observation 4 *N-type countries' incentives to become members of the coalition decline with rising permit prices, while the opposite is true for S-type countries, i.e.,* $\frac{d\phi_N}{dp} < 0$ *and* $\frac{d\phi_S}{dp} > 0$.

Proof: See Appendix. ■

We can now use the above observations to examine the incentives for coalition membership, which allows us to determine the size and composition of the stable coalitions. In particular, the following proposition establishes that the incentives for one type of country depend on the participation of countries of the opposite type.

Proposition 1 *As a higher share of N-type (S-type) countries in the coalition — i.e., a higher (lower) x — raises (lowers) the carbon price, it decreases (increases) the incentives for N-type countries to join the coalition, but raises (lowers) the incentives for S-type countries.*

Proof: See Appendix. ■

The central insight provided by Proposition 1 is that there is *complementarity between N-type and S-type countries* with regards to coalition membership: the incentives for each type of country to join the coalition are negatively affected by a higher share of countries of the same type in the coalition; however, they are positively affected by a higher share of countries of the opposite type.

Coalition Size and Stability

We are now in a position to assess what stable coalitions can arise by examining the stability function ϕ. To start, we simplify the expression for ϕ (14′)

by rewriting it for both types of countries:

$$\phi_N = \frac{1}{2} \cdot p^2/c_N - p \cdot o_N + b_N \cdot o_N - \frac{1}{2} \cdot b_N^2/c_N \tag{14''}$$

$$\phi_S = \frac{1}{2} \cdot p^2/c_S - p \cdot o_S + b_S \cdot o_S - \frac{1}{2} \cdot b_S^2/c_S \tag{14'''}$$

Coalition stability requires that a coalition is internally stable (cf. Carraro and Siniscalco, 1993), meaning no member should have an incentive to leave the coalition, and externally stable, meaning no non-member should have an incentive to join.[11] Here, a coalition is internally stable if neither N-type nor S-type countries have an incentive to leave the coalition, i.e., all members derive a higher net payoff from their membership than from free-riding. This is given by values of x for which $\phi_N \geq 0 \wedge \phi_S \geq 0$. Coalitions are externally stable with respect to type i if the stability function becomes negative when an additional country of this type joins, formally $\phi_N((n_N + 1)/n_S) < 0$ or $\phi_S(n_N/(n_S + 1)) < 0$, or if all countries of type i are already members, $n_i = N_i$ for $i = N, S$. To prepare for Proposition 2 where we will characterize stable coalitions, we now examine the range of permit prices for which N-type and S-type countries prefer to be members of the coalition instead of engaging in free-riding behavior. Within this range, the coalitions are internally stable.

First, taking the roots of the quadratic Equation (14″) for which N-type countries have an incentive to be members of the coalition (keeping in mind A2) yields the expression for feasible prices for N-types:[12]

$$p_{1,2} \leq (c_N \cdot o_N) \pm (c_N \cdot o_N - b_N). \tag{15}$$

Only the stricter constraint is binding. Due to A2, $o_N c_N > b_N$. Hence, the negative sign yields the stricter conditions, such that the incentive compatibility condition for N-type countries results in:

$$p \leq b_N. \tag{15'}$$

---

[11] Note that due to Observation 4, no constellations exist in which all coalition members would be better off by excluding a country from the coalition. There cannot be consensus about restricting membership because excluding a country causes the price to either fall or rise. In either case, one country type loses while the other gains. Hence, internal and external stability appropriately characterize stable coalitions in the context of our model.

[12] The directions of inequalities (15) and (16) result from taking into account the monotonicity properties established in Observation 4 when solving the quadratic equations implied by (14′).

The above expression (15′) states that N-type countries will not pay a price for emissions reductions that exceeds its marginal benefits of climate change mitigation.

Second, (14‴) yields the range of permit prices for which S-type countries have an incentive to be a member of the coalition:

$$p_{3,4} \geq (c_S \cdot o_S) \pm (c_S \cdot o_S - b_S). \tag{16}$$

Due to A2, $o_S c_S > b_S$ such that the positive part of the second term yields the stricter inequality. The resulting incentive compatibility condition for S-type countries is:

$$p \geq 2 \cdot (c_S \cdot o_S) - b_S. \tag{16′}$$

The rationales for conditions (15′) and (16′) are the following: First, (15′) shows that by joining the coalition, an N-type country increases its total amount of abatement (as its reduction commitment lies above its business-as-usual level). As any increase in total abatement yields equal marginal benefits b_N, it will be rational for N-type countries to join the coalition as long as the carbon price (which determines the marginal costs of fulfilling their reduction commitment o_N) does not exceed their marginal benefits. Second, (16′) is derived from the fact that for S-type countries, meeting their higher reduction commitments o_S as members of the coalition results in additional costs that exceed their additional benefits (as in A2 $o_S c_S > b_S$). Thus, they will only become members of the coalition if the carbon price is high enough to provide net revenues from emissions trading that can compensate them for these costs. This relationship is displayed in Figure 1.

Hence, a coalition containing N-type and S-type countries is only stable if, on the one hand, the permit price is low enough to make it attractive for N-type countries to join (15′), but on the other hand, is high enough to make participation worthwhile for S-type countries (16′). Combining (15′) and (16′) directly results in the range of permit prices for which the coalition exhibits internal stability, i.e., the combined incentive compatibility condition:

$$b_N \geq p \geq 2 \cdot (c_S \cdot o_S) - b_S. \tag{17}$$

As a consequence, $b_N \geq 2 \cdot (c_S \cdot o_S) - b_S$ constitutes a necessary, albeit not a sufficient condition (because p depends on a number of parameters, such as o_N) for the existence of a stable coalition. At first sight, it might seem

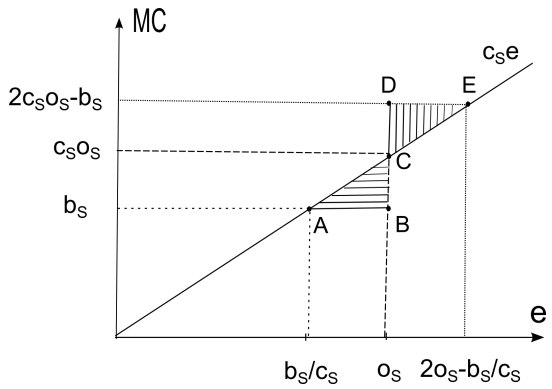

Figure 1. S-type countries' incentives to join the coalition. In the business-as-usual case, S-types abate b_S/c_S. Meeting their higher reduction commitment o_S as members of the coalition results in additional net costs that correspond to the area of triangle ABC. Hence, they will only become members of the coalition if the carbon price is high enough to provide net gains from emissions trading (given by area CDE), which compensate them for these additional costs. Due to the linearity of the marginal cost curve, this condition is fulfilled by a carbon price that is equal to or greater than $c_S o_S + (c_S o_S - b_S)$.

surprising that the participation constraints do not explicitly include the reduction commitment of N-types o_N. It does not enter (17) directly due to the fact that according to (15′) the price p can never rise above b_N, regardless of the number of N-type countries in the coalition or their reduction commitment. In our framework, N-types will never have an incentive to pay a price for emission reductions that exceeds their marginal benefits. This upper limit for p — and hence, for the revenues that S-type countries can generate from selling emissions permits — also implicitly defines the upper limit for S-type countries' reduction commitments o_S because it limits the revenues that S-type countries can derive from selling emissions permits.

However, the highest reduction commitment o_N that is acceptable for N-type countries also depends on the supply of low-cost abatement, and hence the share of S-types in the coalition.[13] In other words, the actual amount of o_N does not matter as long as the price to buy permits is low enough.

[13] The asymmetry that o_S only depends on the parameters, while o_N also depends on the share of S-type countries in the coalition arises because of the assumption of linear benefits (which

Nevertheless, according to (11), p is positively related to o_N such that too high values of o_N will indirectly result in a price that violates (17).

Inserting the incentive compatibility conditions (15′) and (16′) into the expression for p (11′) enables us to rewrite the participation constraints with regard to the ratio of N-type and S-type countries in the coalition:

$$x \leq x_{\max} = \frac{c_N(b_N - o_S c_S)}{c_S(o_N c_N - b_N)} \quad (> 0, \text{ by A2 and A3}) \text{ and} \tag{18}$$

$$x \geq x_{\min} = \frac{c_N(o_S c_S - b_S)}{c_S(o_N c_N + b_S - 2o_S c_S)} \quad (> 0, \text{ by A2 and A3}) \tag{19}$$

Stable coalitions exist when (18) and (19) are simultaneously satisfied, i.e., when there are values of x ($1/N_S \leq x \leq N_N$) that meet both conditions such that at least one country of each type will be a member of the coalition.

The participation constraints (18) and (19) — which determine internal stability — combined with the definition of external stability now allow us to determine the size of stable coalitions in the following proposition.

Proposition 2 *If $x_{\min} \leq N_N/N_S \leq x_{\max}$, the grand coalition will be stable. If $x_{\min} \leq x_{\max} \leq N_N/N_S$, a coalition including all S-type countries will be stable if $\exists n_N \in [1, N_N]$ such that $x_{\min} \leq n_N/N_S \leq x_{\max}$. Likewise, if $N_N/N_S \leq x_{\min} \leq x_{\max}$, a coalition including all N-type countries will be stable if $\exists n_S \in [1, N_S]$ such that $x_{\min} \leq N_N/n_S \leq x_{\max}$.*

Proof: See Appendix. ■

The above proposition highlights one of the central arguments of this paper: widespread participation, and even universal participation in a global climate agreement is feasible with asymmetric countries, emissions trading, and pre-determined reduction commitments. By interpreting the financial transfers that occur through emissions trading as a side payment, this result differs from previous findings for asymmetric countries with joint-welfare maximization, namely that ''allowing for side payments when all countries choose simultaneously to be a signatory or non-signatory does not buy any additional cooperation for the world'' (Barrett, 2001: 1845). The reason for this observation is that without joint-welfare maximization (as in our

determines N-type countries' willingness to pay) and quadratic abatement costs (which determine the supply of abatement by S-type countries).

model), free-rider incentives do not increase with a larger number of coalition members (as is the case in Barrett's model).

Abatement and Payoffs

In this section, we discuss the abatement achieved by stable coalitions relative to the socially optimal abatement level and explore the trade-off between the equitable allocation of emissions permits and achieving the highest possible net payoff.

When coalitions maximize joint welfare, universal participation in the coalition guarantees socially optimal climate change mitigation. Since we depart from this assumption, even the grand coalition will generally not achieve the social optimum. Likewise, it is not a priori clear that stable coalitions would overcome the collective-action problem of public good provision any more than the business-as-usual case would. The following propositions relate the abatement of the coalition to business-as-usual (Proposition 3) and the social optimum (Proposition 4).

Proposition 3 *The maximum additional abatement that can be achieved with a stable coalition is* $e_{tot}^{C} - e_{tot}^{BAU} = N_S \cdot (b_N - b_S)/c_S$*, compared to the business-as-usual case. Abatement is greater if (i) the total number of S-type countries is larger, (ii) the abatement costs of S-types are lower, and (iii) the difference between the benefits of N-type and S-type countries is larger.*

Proof: As a consequence of (9), for any given coalition, maximum abatement occurs if the carbon price is at the maximum level with respect to the (combined) incentive compatibility condition (17), i.e., $p = b_N$. As a coalition member's reduction commitment exceeds that of a non-member (A2), the maximum abatement that can be achieved occurs in a grand coalition with a price of $p = b_N$. Overall abatement is then $e_{tot}^{C} = N_N \cdot (b_N/c_N) + N_S \cdot (b_N/c_S)$, compared to $e_{tot}^{BAU} = N_N \cdot (b_N/c_N) + N_S \cdot (b_S/c_S)$ in the business-as-usual case. The maximum *additional* abatement achievable by cooperation, then, amounts to $e_{tot}^{C} - e_{tot}^{BAU} = N_S \cdot (b_N - b_S)/c_S$.[14] ■

[14] Note that for case (ii), there can also be stable coalitions that do not include all members of any type if x_{min} is 'sufficiently close' to x_{max}.

While Proposition 3 highlights that the coalition achieves higher abatement levels compared to the business-as-usual case, the following proposition evaluates how it performs compared to the socially optimal outcome.

Proposition 4 *Abatement levels that are potentially achievable with full cooperation fall short of the social optimum. The difference between potentially achievable and optimal abatement is greater if (i) the number of countries of each type is larger, (ii) their respective benefits are larger, and (iii) their abatement costs are lower.*

Proof: As demonstrated in Proposition 3, the maximum abatement that can be achieved by a (grand) coalition is $e_{tot}^{C} = (N_N/c_N + N_S/c_S) \cdot b_N$, but according to (8), the socially optimal level would be $e_{tot}^{opt} = (N_N/c_N + N_S/c_S) \cdot (b_N N_N + b_S N_S)$. Hence, the maximum amount of climate change mitigation that is achievable with cooperation falls short of the social optimum by $e_{tot}^{opt} - e_{tot}^{C} = (N_N/c_N + N_S/c_S) \cdot (b_N(N_N - 1) + b_S N_S)$. ■

The collective action problem is magnified by larger numbers of countries, larger benefits and lower costs, and coalitions based on self-interest can achieve relatively less. Thus, while coalitions potentially improve upon the business-as-usual case (according to Proposition 3), Proposition 4 implies that the fundamental collective-action problem cannot be overcome by emissions permit trading alone. Introducing emissions trading gives countries with high benefits access to mitigation options in countries with low mitigation costs, such that they undertake more abatement than they would if they were to rely exclusively on domestic abatement. But, since they do not take other countries' welfare into account, the environmental externality is not fully internalized. Hence, abatement falls short of the socially optimal amount, which through (8), is given by

$$e_{tot}^{opt} = (N_N/c_N + N_S/c_S) \cdot (b_N N_N + b_S N_S).$$

While the proposed agreement would fall considerably short of the social optimum in a world with a large number of countries, it might constitute a viable framework for negotiations that include a smaller number of actors. This would be the case for a regional climate agreement or motions to conclude an international treaty focused on a small number of major emitters, which Victor (2011) proposed as an alternative to the current structure of negotiations.

From the results so far, the specific role of the exogenous reduction commitments is not obvious. In particular, could an agreement where countries engage in emissions trading without joining a coalition (as in Helm, 2003) achieve outcomes similar to those described above? We compare our setting with predetermined reduction commitments to the alternative setting of emissions trading and freely chosen abatement levels to isolate the effects of exogenous reduction commitments and emissions permit trade. The following proposition shows how total emissions reductions can be decomposed into the effect of emissions trading and the effect of predetermined reduction commitments.

Proposition 5 *Two effects contribute to the maximum additional abatement compared to the business-as-usual case: (i) the effect of introducing emissions trading and (ii) the effect of predetermined reduction commitments. Of these two, predetermined reduction commitments make the most important contribution. Their effect on abatement is unambiguously positive and is always stronger than the effect of trading emissions, which can be positive or negative.*

Proof: See Appendix. ■

This highlights the importance of predetermined reduction commitments as part of the proposed agreement studied in this paper, because the predetermined reduction commitments create the conditions for an agreement that can achieve more than emissions trading alone and can guarantee that nontrivial (positive) emission reductions are achieved.

The decisive role of reduction commitments for the maximum abatement of coalitions also translates to the stability of coalitions. In particular, in the following observation, we summarize how countries' incentives to become members of the coalition depend on their abatement obligation:

Observation 5 *The coalition's stability crucially depends on the reduction commitments o_N and o_S that are allocated to N-type and S-type countries, respectively. In particular, a grand coalition can be obtained through the appropriate selection of reduction commitments.*

Proof: See Appendix. ■

Previous studies that examine different allocation rules for emissions permits (such as grandfathering, equal-per-capita, or contraction and convergence) find only modest increases in coalition size and global abatement (e.g., Altamirano-Cabrera and Finus, 2006), while others arrive at more optimistic results (Carraro *et al.*, 2006; Weikard, 2009; Nagashima *et al.*, 2009). Observation 5 emphasizes that with an appropriate sharing rule, significant improvements in participation can be achieved, but not all allocation schemes are optimal. As observed in Proposition 3, a higher level of participation will also result in more abatement, but how much of the gap to the social optimum can be closed depends on the parameters (see Proposition 4). In particular, the joint incentive compatibility constraint, (17), shows that the highest permit price for both types of countries to have an incentive to be members of the coalition is $p = b_N$. As shown in Proposition 3, if reduction commitments are allocated in a way such that a stable coalition forms at this price, it may be a grand coalition. Using (11), it is easily shown that the best achievable outcome in terms of abatement can be achieved by the following allocation of emissions permits:

$$o_S = (b_N + b_S)/2c_s \tag{20}$$

$$o_N = b_N/c_N + N_S/N_N \cdot (b_N - b_S)/2c_s \tag{21}$$

Besides demonstrating the importance of the distribution of reduction commitments, the observation also has important implications for climate policy. Universal participation in a global climate agreement can be achieved through the adequate selection of reduction commitments, which put an upper limit on the overall level of mitigation that can be achieved. However, nothing guarantees that such a distribution is in accordance with fundamental equity considerations, such as distributing emissions permits on an equal per-capita basis or based on historical responsibility (see Markandya, 2011, for an overview of the relevant equity dimensions and Bodansky, 2004, for a summary of the proposed allocation principles). This observation is in line with Germain and van Steenberghe (2003) who point out that is unlikely that most equitable allocation rules are individually rational for countries that would be required to bear relatively large shares of the mitigation burden.

Finally, we examine the coalition size and stability of a coalition that aims to maximize the joint welfare of its members instead of predetermined reduction commitments.

Proposition 6 *If the coalition aims for maximum joint welfare instead of relying on exogenously given reduction commitments, no stable coalition can form.*

Proof: See Appendix. ■

No coalition is stable under joint-welfare maximization because the price that would maximize the coalition's welfare violates the participation constraint of N-type countries. This result is in line with findings by Fuentes-Albero and Rubio (2010) who present a standard model in which symmetric countries can form a stable coalition of three members, while there is no stable coalition with asymmetric benefits from abatement and only a stable coalition of two countries with asymmetric abatement costs. Asymmetry without additional mechanisms, such as emissions trading or transfer schemes, makes cooperation more complicated under the assumption of joint-welfare maximization because the different interests of the member countries are not reconciled.

Discussion and Conclusions

The literature on coalition formation has repeatedly emphasized that self-interested behavior produces strong incentives for free-riding. A high level of cooperation is then unlikely to occur. The model presented in this paper shows how emissions trading, in combination with a predetermined allocation of emissions permits, can exploit countries' self-interest and yield a higher payoff for every country and more overall abatement, compared to the business-as-usual case. Our analysis shows that, while emissions trading in conjunction with an appropriate allocation of emissions permits creates an incentive to join the coalition by distributing the economic surplus generated by equalizing the marginal abatement costs across countries, it does not solve the underlying collective action problem. That is, the resulting outcome falls short of the social optimum, even when full participation is achieved. Furthermore, allocation schemes that guarantee that all countries have an incentive to join the coalition might turn out to be fundamentally at odds with equity considerations, such as distributing emissions permits on an equal per-capita basis or based on historical responsibility for the already existing stock of greenhouse gases in the atmosphere due to past emissions.

From a policy perspective, our results suggest that there is an advantage to ''packaged deals'' that bundle the participation decision with reduction commitments specified by fixed rules that are not subject to negotiation. We argue that this can help to achieve broader participation in a climate agreement. These rules have to be designed in a way that makes it individually rational for each country to participate — i.e., they may be required to strike a balance between pragmatism and equity considerations. Furthermore, our analysis suggests that the resulting agreement will be broad but shallow, meaning each country's reduction commitments will fall short of the socially optimal level. Thus, while such an agreement can be an improvement over the business-as-usual outcome, it generally cannot solve the underlying collective action problem. Other mechanisms will be required in order to achieve true cooperation in which all of the external effects of greenhouse gas emissions are fully internalized.

References

Altamirano-Cabrera, J. C. and M. Finus. 2006. "Permit Trading and Stability of International Climate Agreements." *Journal of Applied Economics* 9(1): 19–48.

Altamirano-Cabrera, J. C., M. Finus, and R. Dellink. 2008. "Do Abatement Quotas Lead To More Successful Climate Coalitions?" *Manchester School* 76(1): 104–129.

Barrett, S. 1994. "Self-Enforcing International Environmental Agreements." *Oxford Economic Papers, New Series*, Vol. 46, Special Issue on Environmental Economics, pp. 878–894.

Barrett, S. 1997. "Heterogenous International Environmental Agreements." In *International Environmental Negotiations, Strategic Policy Issues*, C. Carraro, ed., Cheltenham: Edward Elgar.

Barrett, S. 2001. "International Cooperation for Sale." *European Economic Review* 45(10): 1835–1850.

Barrett, S. 2002. "Consensus Treaties." *Journal of Institutional and Theoretical Economics* 158(4): 529–547.

Biancardi, M. and G. Villani. 2010. "International Environmental Agreements with Asymmetric Countries." *Computational Economics* 36: 69–92.

Bodansky, D. 2004. "International Climate Efforts Beyond 2012: A Survey of Approaches." Prepared for the Pew Center on Global Climate Change.

Bosello, F., B. Buchner, and C. Carraro. 2003. "Equity, Development, and Climate Change Control," *Journal of the European Economic Association* 1(2–3): 601–611.

Botteon, M. and C. Carraro. 1997. "Burden-Sharing and Coalition Stability in Environmental Negotiations with Asymmetric Countries." In: *International Environmental Negotiations, Strategic Policy Issues*, C. Carraro, ed., Cheltenham: Edward Elgar, pp. 26–55.

Carbone, J., C. Helm, and T. F. Rutherford. 2009. "The Case for International Emission Trade in the Absence of Cooperative Climate Policy." *Journal of Environmental Economics and Management* 58(3): 266–280.

Carraro, C., J. Eyckmans, and M. Finus. 2006. "Optimal Transfers and Participation Decisions in International Environmental Agreements." *The Review of International Organizations* 1(4): 379–396.

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

Colmer, J. 2011. "Asymmetry, Optimal Transfers and International Environmental Agreements." Prepared for the Grantham Research Institute on Climate Change.

Eyckmans, J. and M. Finus. 2006. "Coalition Formation in a Global Warming Game: How the Design of Protocols Affects the Success of Environmental Treaty-Making." *Natural Resource Modeling* 19(3): 323–358.

Finus, M. 2003. "Stability and Design of International Environmental Agreements: The Case of Transboundary Pollution." In *International Yearbook of Environmental and Resource Economics 2003/4*, H. Folmer and T. Tietenberg, eds., Cheltenham: Edward Elgar.

Finus, M. 2008. "Game Theoretic Research on the Design of International Environmental Agreements: Insights, Critical Remarks, and Future Challenges." *International Review of Environmental and Resource Economics* 2: 29–67.

Finus, M. and S. Maus. 2008. "Modesty May Pay!" *Journal of Public Economic Theory* 10(5): 801–826.

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

Germain, M. and V. van Steenberghe. 2003. "Constraining Equitable Allocations of Tradable CO_2 Emission Quotas by Acceptability." *Environmental & Resource Economics* 26(3): 469–492.

Hannesson, R. 2010. "The Coalition of the Willing: Effect of Country Diversity in an Environmental Treaty Game." *Review of International Organization* 5: 461–474.

Helm, C. 2003. "International Emissions Trading with Endogenous Allowance Choices." *Journal of Public Economics* 87(12): 2737–2747.

Hoel, M. 1992. "International Environment Conventions: The Case of Uniform Reductions of Emissions." *Environmental and Resource Economics, Springer* 2: 141–159.

IPCC. 2007. "Climate Change 2007. Mitigation of Climate Change." In *Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, B. Metz, O. R. Davidson, P. R. Bosch, R. Dave, and L. A. Meyer, eds., Cambridge: Cambridge University Press.

Jaeger, C. C. and J. Jaeger. 2011. "Three Views of Two Degrees." *Regional Environmental Change* 11: 15–26.

Markandya, A. 2011. "Equity and Distributional Implications of Climate Change." *World Development* 39(6): 1051–1060.

McGinty, M. 2007. "International Environmental Agreements Among Asymmetric Nations." *Oxford Economic Papers* 59(1): 45–62.

Nagashima, M., R. Dellink, E. van Ierland, and H. P. Weikard. 2009. "Stability of International Climate Coalitions — A Comparison of Transfer Schemes." *Ecological Economics* 68: 1776–1787.

Nordhaus, W. 2008. *A Question of Balance*. Yale University Press.

Victor, D. G. 2011. *Global Warming Gridlock: Creating More Effective Strategies for Protecting the Planet*. Cambridge University Press.

Weikard, H. P. 2009. "Cartel Stability Under An Optimal Sharing Rule." *Manchester School* 77(5): 575–593.

Weikard, H. P. and R. Dellink. 2010. Sticks and carrots for the design of international climate agreements with renegotiations. Annals of Operations Research, 1-20.

Weikard, H. P., M. Finus, and J. C. Altamirano-Cabrera. 2006. "The Impact of Surplus Sharing on the Stability of International Climate Agreements." *Oxford Economic Papers* 58: 209–232.

Appendix: Proofs

Proof of Observation 1

Observation 1 states that $p(x)$ (strictly) increases in the ratio $x = n_N/n_S$. To see this, consider the derivative of $p(x)$, x to be real. As in accordance with $(11')$ $\frac{dp(x)}{dx} > 0$, $p(x)$ strictly increases, which carries over when $p(x)$ is restricted to the discrete domain of $p(n_N/n_S)$. The carbon price, therefore, strictly increases in x, or $\frac{\Delta p}{\Delta x} > 0$. ■

Proof of Observation 2

Observation 2 follows directly from calculating the limits of $(11')$ for $x \to 0$ and $x \to \infty$ in combination with A3 and Observation 1. Similar to the proof of Observation 1, we can only consider the limits of the continuous function $p(x)$. Obviously, the upper and lower bounds found in this way also constrain $p(n_N/n_S)$. ■

Proof of Observation 3

Observation 3 follows directly from combining (9) with Observation 2, which yields $e_N^C - o_N = p/c_N - o_N < 0$ and $e_S^C - o_S = p/c_S - o_S < 0$. ■

Proof of Observation 4

Observation 4 is rather intuitive, given that N-type countries are net importers and S-type countries are net exporters of emission permits, as established in Observation 3. Formally, it can easily be shown that $\frac{d\phi_i}{dp} = p/c_i - o_i$ $(i = \{N; S\})$, which in combination with Observation 2, yields $\frac{d\phi_N}{dp} < 0$ and $\frac{d\phi_S}{dp} > 0$. ■

Proof of Observation 5

First, the (combined) participation constraint, (18), establishes an upper limit for the maximum reduction commitment for which S-type countries have an incentive to join the coalition: $o_S \leq (b_N + b_S)/2c_S$. Second,

(19), in combination with the condition that $x \geq 1/N_S$, results in the upper limit for the reduction commitments of N-type countries: $o_N \leq b_N(1/c_N + N_S/c_S) - o_S$. Hence, stable coalitions that satisfy both participation constraints can be obtained by choosing the appropriate reduction commitments, o_N and o_S. More specifically, noting that for reduction commitments sufficiently close to business-as-usual levels of abatement, i.e., $o_S \to b_S/c_S$, $x_{\min} \to 0$ and for $o_N \to b_N/c_N$, $x_{\max} \to \infty$, a coalition featuring full membership can be obtained by appropriate allocation of reduction commitments. Yet, this does not mean that any desired level of abatement can be achieved by choosing reduction commitments accordingly; rather, as shown in Proposition 4, total abatement is strictly below the social optimum. ■

Proof of Proposition 1

Proposition 1 follows directly from the monotonicity properties of $\phi_i(p)$ and $p(x)$ established in Observations 1 and 4, which can be combined to yield $\frac{\Delta\phi_N}{\Delta x} < 0$ and $\frac{\Delta\phi_S}{\Delta x} > 0$. ■

Proof of Proposition 2

Proposition 1 has shown that $\frac{\Delta\phi_N}{\Delta x} < 0$ and $\frac{\Delta\phi_S}{\Delta x} > 0$. This general behavior of the stability functions ϕ_N and ϕ_S is sketched in Figure A1. Due to the discrete nature of n_N and n_S , and therefore $x = n_N/n_S$, $\phi_i(x)$ only takes on discrete values on the depicted continuous lines.

Recall that coalitions are stable if $\phi_N \geq 0 \wedge \phi_S \geq 0$ and $\phi_N((n_N + 1)/n_S) < 0 \wedge \phi_S((n_N/(n_S + 1)) < 0$, or if all countries of type i are already members, $n_i = N_i, i = \{N, S\}$. That is, coalitions are stable with respect to type N (or S) at the largest x with $\phi_N(x) \geq 0$ and the lowest x with $\phi_S(x) \geq 0$, respectively. In Figure A1, this happens at the intercepts of the stability function (i.e., $\phi_N = 0$ and $\phi_S = 0$). In sum, a coalition is stable if one of the following conditions holds for both types, i.e., either x is at the intercept of this country type's stability function, or the stability function is non-negative at x and participation of this type of country is full.

For $x_{\min} > x_{\max}$, as defined in (18) and (19) and depicted in panel (a), internal stability fails for all values of x, thus no stable coalition emerges. If, on the other hand, $x_{\min} < x_{\max}$, a stable coalition exists. The size and

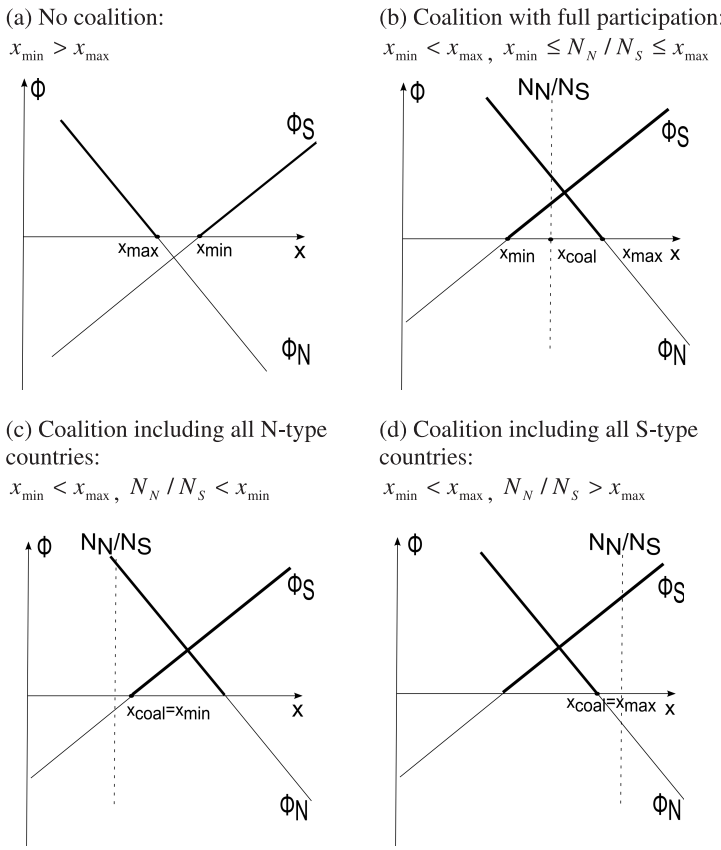

Figure A1. Possible outcomes of the coalition game.

composition of the stable coalition is determined by the complementarity between N- and S-type countries laid out in Proposition 1, which always makes joining the agreement attractive for at least one country type. From any x for which $\phi_i > 0$ ($i = \{N; S\}$), free-riding N- and/or S-type countries would join the coalition until either (i) no non-members are left or (ii) one type has no more incentives to join, that is, one of the participation constraints specified in (18) and (19) is reached, i.e., $x = x_{\min}$ or $x = x_{\max}$.

Case (i) holds if all countries are coalition members and, thus, the contingent of non-members is exhausted before one of the participation constraints is reached, as depicted in Panel (b). That is, for $x_{\min} \leq N_N/N_S \leq x_{\max}$, a coalition with full participation will be stable, meaning no country will have an incentive to leave.

For case (ii) (i.e., if $N_N/N_S < x_{\min}$, or $N_N/N_S > x_{\max}$), countries will join the coalition until all the countries of one type (N-type or S-type, respectively) are members and countries of the other type have no more incentives to join (i.e., one of the participation constraints is reached), provided that there are appropriate integer numbers for both types of countries whose fraction lies inside the interval $[x_{\min}; x_{\max}]$ (i.e., $\exists\, n_N \in [1, N_S]$ such that $x_{\min} \leq n_N/N_S \leq x_{\max}$ or $\exists n_S \in [1, N_S]$ such that $x_{\min} \leq N_N/n_S \leq x_{\max}$, respectively).[15] Panel (c) illustrates the case in which external stability is reached with a value of zero for the stability function for S-type countries (i.e., S-type countries that remain outside the coalition have no incentive to join), while for N-type countries, the stability function is positive at $x = x_{\min}$. That is, external stability is obtained with all N-type countries being coalition members. Coalition membership is consequently determined by $\{n_N = N_N; n_S = N_N/x_{\min}\}$. Likewise, Panel (d) shows the case in which external stability is reached by a zero value for the stability function of N-type countries, and all S-type countries are members of the coalition (as at $x = x_{\max}$ where their stability function is positive). Coalition membership is then determined by $\{n_N = N_S \cdot x_{\max}; n_S = N_S\}$. ■

Proof of Proposition 5:

As a benchmark for our comparison, we examine a setting in which all countries participate in emissions trading and each country can freely choose its reduction target. Due to symmetry, we can then take as given that all N-type countries (or S-type countries) choose identical reduction targets.

The payoff function, (13), depends on the price p and, hence, on both o_N and o_S. Therefore, we can calculate a Nash-equilibrium in reduction targets by maximizing (13) for N-type countries, as well as S-type countries as a best response to the other players' equilibrium strategies. Using X to denote the ratio of N-type countries to S-type countries ($X = N_N/N_S$), we obtain two equations that can be solved for the two unknowns, o_N^* and o_S^*:

$$o_N^* = \frac{b_N}{c_N} + \frac{b_N - b_S}{2c_S \cdot X}$$
$$o_S^* = \frac{b_S}{c_S} + \frac{(b_N - b_S) \cdot X}{2c_N}.$$

[15] Note that for case (ii), there can also be stable coalitions that do not include *all* members of any type of country if x_{min} is "sufficiently close" to x_{max}.

The first summand of these two expressions corresponds with the respective BAU abatement levels (cf. Equation (4)). If the countries were symmetric (i.e., $b_N = b_S$), the second summand would vanish and each country would simply abate as in BAU, and no emissions trading would occur. However, in assumption A1 (i.e., $b_N > b_S$), N-type countries (or S-type countries) choose a higher (or lower) reduction commitment than under the BAU and, hence, they choose a lower (or higher) endowment with emissions permits. This finding mirrors Helm's (2003) observation that in a model with a freely chosen reduction target, ''environmentally more concerned countries usually choose less allowances if these are tradable, but this may be offset by the choice of more allowances on the side of environmentally less concerned countries'' (p. 2737). We can now compare the total abatement that occurs in the case with emissions trading and freely chosen emissions permits ($e^*_{tot} = N_N o^*_N + N_S o^*_S$) with the BAU abatement (e^{BAU}_{tot}), as given by (5):

$$e^*_{tot} - e^*_{BAU} = \frac{(b_N - b_S) \cdot (N_S c_N - N_N c_S)}{2 c_N c_S}.$$

Depending on the parameters, this expression can be positive or negative. That is, since under freely chosen reduction commitments S-type countries choose endowments with emissions permits that are higher than their BAU emissions, total emissions can potentially increase. In such cases, permit trading alone is obviously not sufficient to result in emissions reductions, and needs to be complemented by predetermined reduction commitments, as in our model.

We can now decompose the contribution of emissions trading and predetermined reduction commitments. Noting that the maximum improvement with respect to the BAU identified in Proposition 3 is given by $e^C_{tot} - e^{BAU}_{tot} = N_S(b_N - b_S)/c_S$, we can denote the contribution of emissions trading as

$$\frac{e^*_{tot} - e^{BAU}_{tot}}{e^C_{tot} - e^{BAU}_{tot}} = \frac{N_S c_N - N_N c_S}{2 c_N N_S},$$

and the remaining contribution, which can be attributed to predetermined reduction commitments as:

$$1 - \frac{e^*_{tot} - e^{BAU}_{tot}}{e^C_{tot} - e^{BAU}_{tot}} = \frac{N_S c_N + N_N c_S}{2 c_N N_S}.$$

Whereas the first expression can be negative or positive, the second is strictly positive and greater than the first one. That is, the addition of a predetermined reduction commitment makes a greater contribution towards closing the gap between e_{tot}^{C} and e_{tot}^{BAU} than the introduction of emissions trading without reduction commitments does. ■

Proof of Proposition 6:

In analogy to (7), it is straightforward that for a given coalition of size $\{n_N, n_S\}$, welfare is maximized when each member's marginal abatement costs equal the sum of all the members' marginal benefits. In combination with (9), this yields the familiar condition that the permit price equals the coalition's marginal benefit: $p = n_N b_N + n_S b_S$. This price can be attained by an appropriate choice of $\{o_N, o_S\}$ in (11). However, in (17), N-type countries only have an incentive to be in the coalition as long as $p \leq b_N$. The price that would maximize the coalition members' welfare, thus, violates the incentive for a compatibility condition for N-type countries. Hence, keeping in mind the restriction $n_N > 0$ and $n_S > 0$, no stable coalition is feasible under joint-welfare maximization. ■