Games **2010**, *1*, 137-158; doi:10.3390/g1020137

OPEN ACCESS

games

ISSN 2073-4336

www.mdpi.com/journal/games

Article

Can Justice and Fairness Enlarge International Environmental Agreements?

Christine Grüning and Wolfgang Peters *

Department of Business Administration and Economics, European University Viadrina, PO Box 1786, D-15207 Frankfurt (Oder), Germany

* Author to whom correspondence should be addressed; E-Mail: peters@euv-frankfurt-o.de; Tel. +49-335-55-34-25-90; Fax +49-335-55-34-22-38.

Received: 11 May 2010; in revised form: 1 June 2010 / Accepted: 14 June 2010 / Published: 24 June 2010

Abstract: The literature on International Environmental Agreements (IEAs) predicts a rather low number of signatories to an IEA. This is in sharp contrast to empirical evidence. As experimental economics provides some evidence for more complex human behavior, extending the theory of IEAs to a broader class of preferences is clearly promising. The present paper shows that where countries' preferences incorporate justice and fairness there will be a strong incentive for them to choose similar abatement policies within and outside an IEA. Consequently, free-riding at the expense of the signatory states diminishes and participation in an IEA becomes a more successful strategy, so that the size of stable IEAs increases.

Keywords: International Environmental Agreements; coalition formation; justice and fairness

---

1. Motivation

Where supra-national institutions are absent voluntary cooperation between countries is designed to restrict harmful impacts of global environmental problems like greenhouse gas emissions which cause a shrinking ozone layer and global warming. Although not all countries cooperate, we nevertheless observe several International Environmental Agreements (IEAs) where the number of signatories is decidedly large. However, even with an IEA cooperation suffers from free-riding among the remaining

non-signatories and between insider and outsider countries such that they do not go far beyond 'business as usual'. Thus, effectiveness of an IEA needs not significantly be improved with the coalition size. Murdoch and Sandler [1] and more recently Böhringer and Vogt [2] state that there do exist agreements with a huge number of signatories and a rather lax abatement level similar to the non-cooperative Nash outcome, e.g., the Kyoto Protocol.[1] Although (and maybe because) the overall environmental policy with and without international agreements are rather similar, we do observe huge participation and very weak agreements at the same time. However, this empirical evidence is in sharp contrast to theoretical prediction. It is a well-known result in the theory on IEAs that the number of countries actively engaged in an IEA is likely to be very small.[2] To reconcile empirical evidence and economic theory the standard literature must be extended so that the huge participation in International Climate Change Agreements can be explained.[3]

Standard IEA models assume that each country is concerned only with its own welfare, and exclusively address the difference between environmental benefits of a country and its abatement costs. However, there are some hints from experimental economics that human behavior is more complex than pure selfishness.[4] Among others Ringius *et al.* [20] identify fairness as an important motivation for countries' behavior in environmental negotiations. They analyze actual IEAs and trace some of the agreement characteristics back to an underlying principle of justice between participants. Additionally, an empirical analysis by Lange *et al.* [21] focuses on fairness ('equity considerations') during the process of negotiations about international climate agreements. Almost all interviewed experts involved in the Kyoto negotiation process state that fairness plays an important role.[5] Accordingly, justice and fairness must be addressed in IEA models as additional motive for governments' behavior.

Fairness puts some pressure on governments to accept similar responsibilities.[6] Thus, countries apply relatively conforming measures or strategies which can be implemented in the theoretical analysis either by new instruments or by extending governments' objectives. The former strategy was used by Hoel [8] or Finus and Rundshagen [28] who focus on uniform emission reductions or uniform quotas. Here, an institutional restriction obliges countries to behave alike but there is still no endogenous motivation for

[1] From a political economy perspective Böhringer and Vogt [3] argue that the rather poor results of the Kyoto Protocol can best be understood as a stance of a symbolic policy.

[2] For details see Barrett [4,5], Carraro and Siniscalco [6], Finus [7], or Hoel [8].

[3] See Barrett [9], Barrett and Stavins [10] or Buchholz and Peters [11] for recent approaches to reinforce the incentives to engage in IEAs. A strong recommendation for IEAs is given by Stern [12].

[4] Cf. Alesina and Angeletos [13], Bolton and Ockenfels [14], Falk *et al.* [15], Fehr and Schmidt [16], or Rabin [17] and the literature cited in these papers. For a more general view on interdependent preferences and reciprocity, see Sobel [18]. Introducing justice and fairness bears the risk of modelling the extended preferences arbitrarily, as a broad range of observations can be used to explain any type of behavior, cf. Postlewaite [19].

[5] Lange *et al.* [21] asked in their questionnaire for both, experts' own view on equity as well as the perception of equity views in different countries or country groups. Nevertheless, behind the surface these fairness considerations are often driven by a material self interest, cf. Lange *et al.* [22].

[6] According to Albin [23] fairness has multiple facets (e.g., altruism or reciprocity) and is not unambiguously defined. Both concepts put some pressure on governments to behave conform and thus enforce behavioral or social norms. Following Lindbeck [24] norms have an impact on rational behavior. As shown in Wooders *et al.* [25] self-interested behavior, conformity and social norms need not be inconsistent. This is in line with Elster [26], p. 102 who noticed that individual "actions typically are influenced both by rationality and by norms". For more details about social norms and private provision of public goods see Rege [27].

such an institutional rule. Hence, extending the class of preferences aims at an endogenous explanation of conforming behavior.

Hoel and Schneider [29], followed by Jeppesen and Andersen [30], first applied a broader class of preferences to IEAs. They assume that countries are not only concerned with their own welfare. Countries' well-being is also related to the behavior of the other countries, so that becoming a member of an IEA is an end in itself. As countries in this setting have a bias to join an agreement, governments exhibit fairly conforming attitudes. In a recent approach Lange and Vogt [31] introduced a self-centered equity concern as an additional motivation for cooperation by applying Bolton and Ockenfels' [14] ERC preferences.[7] Governments with ERC preferences prefer *an own* behavior close to the average. Lange and Vogt [31] focus on countries' welfare which is unobservable. However, Lange *et al.* [21] empirically address governments' perceptions of equity and fairness based on Ringius *et al.* [20] and their implementation in IEA negotiations. They found that countries primarily focus on observable abatements so that free-riding at the expense of others conflicts with fairness.[8] Summarizing, countries with fairness-oriented preferences compare their own measures or abatement costs with that of all other countries. All countries will contribute when they suppose that all (or sufficiently many of them) will comply as well. Furthermore, countries dislike own and other countries' deviations from this conforming behavior.[9] This is more in line with fairness-oriented preferences identified by Fehr and Schmidt [16]. They assume that countries dislike a deviation of *all* countries from the average, *i.e.,* prefer an equal distribution of abatement cost among countries. Thus, a fair cost sharing of abatement duties results.[10] A situation where the distribution of abatement costs is relatively heterogeneous among countries will be regarded as unfair.[11] "People ... appear to desire equality relative to some reference point, namely what they consider to be 'fair' payoffs".[12] Thus, governments tend to avoid cost dispersion that can be measured in a basic approach by the variance of abatement costs.[13] In this symmetric case, disutility is the same whether a country abates *more* or *less* than the average. However, experimental data from Blanco *et al.* [35] and Dannberg *et al.* [36] show that there is at least some asymmetry regarding both kinds of inequality. Thus, in an extension to our basic approach we do allow for preferences which treat these deviations differently.

Aside from fairness-oriented preferences, we also integrate two types of offsetting behavior. As we will show subsequently, complete or partial free-riding (doing nothing or applying a moderate policy

[7] ERC stands for equity, reciprocity and competition.

[8] See Victor and Coben [32] for a different view of countries' conform behavior. They suggest that quantity strategies are favored by the diplomatic community over price instruments. While equal treatment can easily be granted with the former instrument, the latter often results in heterogeneous quantity effects among countries.

[9] This point of view directly corresponds to Rawls' [33], p. 236 theory of justice.

[10] According to Engelmann and Strobel [34] countries with Fehr and Schmidt [16] preferences are superior to ERC preferences in explaining observations from experiments.

[11] There is no doubt about this statement as long as countries are symmetric or rather similar. But in real life, countries differ with respect to important variables. In case of heterogeneity, the principle of equal treatment requires that different agents must be treated differently. In an extension to our basic approach, we allow for heterogeneous countries and preferences.

[12] Alesina and Angeletos [13], p. 965.

[13] Alesina and Angeletos [13] apply the variance as an appropriate measure for fairness and justice. Although it seems to be obvious that a player suffers more from an inequality that is to his disadvantage, we assume—for simplicity reasons—that countries attribute a similar or the same disadvantage if they abate less or more than the others. Unless the level of abatements, coalition size and threshold for partial free-riding vary slightly, both assumptions yield to qualitatively similar results.

only) plays a crucial role.[14] The larger the number of IEA members, the more the signatories internalize the global externality. Accordingly, the incentive to provide additional measures outside the coalition diminishes, and any non-signatory state becomes a complete free-rider. Therefore, integrating both types of free-riding allows to study the offsetting behavior outside the IEA appropriately.

The economic intuition for coalition formation distinguishes between three underlying motives: the traditional ones, *i.e.,* countries' *individual gain from free-riding*, and the *collective efficiency gain* which measures the internalization of the environmental externality and in addition the impact of justice and fairness, which favors similar behavior with respect to abatements (*conforming behavior*). The last aim can best be met when each country has a similar participation strategy. This results in either the grand coalition or complete failure of the negotiation process. As the traditional effects work in opposite directions and fairness destabilizes medium-sized coalitions, it is the interplay of all three effects that determines the equilibrium of the entire game.

The paper is organized as follows. In section 2 we present the economic framework. Subsequently, in section 3 we analyze the policy game on abatements, which is followed by the formation of an IEA through a coalition of countries. Section 5 discusses an extension to our basic approach allowing for asymmetric preferences and some heterogeneity. Finally, we present some concluding remarks.

2. Economic Framework

In what follows we study in a complete information world a standard coalition formation game like that introduced by Barrett [4,5] or Carraro and Siniscalco [6]. The aim is to explain international cooperation for $N \geq 4$ identical countries in case of an IEA.[15] In a first stage, countries can choose whether or not to join an IEA. This decision process results in S signatory states with the remaining $(N - S)$ countries behaving non-cooperatively. Subsequently, in stage two, insider and outsider of an IEA decide simultaneously on their abatement measures.[16]

The choices at both stages are determined through rational behavior of all countries. Instead of following the traditional approach, which focuses on the net benefit of global abatement strategies and private costs of the environmental policy, we additionally rely on preferences which directly integrate fairness considerations. As governments can easily observe countries' policies, they can see whether

[14] Most approaches analyzing coalition formation on IEAs stick either to the case where all non-signatories are complete free-rider or they assume that even outsiders do not completely refrain from environmental measures, cf. Lange and Vogt [31]. Nevertheless, free-riding depends at least on the number of already cooperating countries within an IEA. Thus, complete free-riding is not exogenous and should, therefore, be analyzed endogenously.

[15] As we tackle global environmental problems, the number of countries involved is sufficiently large and $N \geq 4$ is a rather weak assumption. If there are less than four countries there is no incentive to stay outside an IEA. However, we focus on agreements where the participation is endogenous and should therefore not be predetermined.

[16] Contrary to Barrett [4,5] we simplify the strategic interaction between signatories and singletons by neglecting Stackelberg behavior of the coalition. Thus, we focus on simultaneously acting countries like in Carraro and Siniscalco [6]. However, as we deal with strategic substitutes, the coalition's weak position produces a more engaged IEA. Consequently, the abatement activities of a member and an outsider are more polarized than under the Stackelberg assumption. In contrast, members of an IEA playing a leading role vis-à-vis the non-signatory states can reinforce outsiders' abatements through a reduction in their own activities. This directly favors the member states at the expense of the outsiders. In the case of justice and fairness, Stackelberg leadership stabilizes larger coalitions even more as the economic behavior of insiders and outsiders turn out to be rather homogeneous.

other countries are more engaged in environmental concerns. Then, their own deviations from a homogeneous strategy as well as foreign ones can be seen as a welfare loss. Hence, justice and fairness focus on the dispersion of countries' observable abatement costs. As a consequence, country j's payoff consists of the benefit minus costs[17]—represented by a quasi-linear logarithmic function—minus a term which measures heterogeneity by means of the variance in all abatement strategies

$$P_j = \ln\left(\sum\nolimits_i a_i\right) - a_j - \theta \cdot \sigma\left(a_1, ..., a_N\right) \tag{1}$$

where $(a_i, a_j) \geq 0$ correspond to the abatement levels of country i and j, while $\sigma(a_1, ..., a_N)$ measures the variance in the environmental policy of all countries. According to Alesina and Angeletos [13] the variance is a good measure for fairness and justice, as countries prefer a more egalitarian cost sharing. The variance in the abatement strategies is defined as $\sum_i \frac{1}{N}\left(a_i - \overline{a}\right)^2$, where $\overline{a}$ is the global average of all countries' environmental policies.[18] A country's payoff is strictly concave in its own strategy and continuous in that of the opponents. Moreover, in order to analyze the impact of justice and fairness, we introduce a parameter, $\theta \geq 0$, that represents the preference intensity for the welfare loss due to cost dispersion. Thus, in case of $\theta = 0$, the payoff function of country j coincides with that in the traditional approaches which focus on pure selfishness. Increasing θ corresponds with a stronger concern for 'just or fair' cost sharing.

Following the literature on IEA, we have a two-stage game, where the countries decide at a first stage whether to sign an IEA, given the decision of all other countries. Such a decision has to be based on what countries do after the signatories of the IEA and the outsiders are determined. For this reason, countries need to anticipate the level of abatements of the countries both inside and outside the coalition which will be established at stage two. Furthermore, IEAs are voluntary alliances of at least two countries $(S \geq 2)$. All signatory states S behave cooperatively among themselves, whereas $(N - S)$ singletons behave non-cooperatively towards both the coalition and each other. The voluntary nature of an IEA implies that a country joins a coalition only if this is a reasonable strategy for a potential signatory. Needless to say, equilibrium participation in an IEA requires both internal and external stability, *i.e.,* no insider and no outsider has an incentive to deviate from the chosen participation strategy.[19]

Our principal objective is to analyze whether the number of IEA members is positively correlated with issues of justice and fairness in countries' preferences. However, before studying stability, we solve the game by backward induction.

3. Policy Game: The Second Stage

The countries simultaneously determine their abatement strategies at the second stage of the game. In the presence of a positive global environmental spillover, the voluntary cooperation of some countries improves the situation of the remaining singletons and creates an incentive to free-ride. If countries have identical preferences, signatories are—in equilibrium—more engaged than outsiders $(a_s > a_o)$. As singletons can decide whether, and how to engage in environmental concerns, we distinguish between

[17] For simplicity, we assume that costs are linear in abatements.

[18] According to Rege [27] the global average $\overline{a}$ can be seen as a norm for conform behavior.

[19] This definition corresponds to that of cartel stability presented in the oligopoly literature by d'Aspremont and Gabszewicz [37].

complete $(a_o = 0)$ and partial $(a_o > 0)$ free-riders. Existence of a Nash equilibrium in the abatement game is guaranteed as the payoff function is strictly concave in the own strategy and continuous in the opponents' strategies. Additionally, the equilibrium is unique as we can apply a more general proof for the coalitional equilibria of Finus *et al.* [38] to our framework.[20] Consequently, in accordance with identical preferences and uniqueness, all signatories and outsiders are symmetric.

The cooperatively acting members of an IEA maximize their joint payoff, given the abatement levels of the non-cooperative countries a_o. Due to symmetry, the coalition maximizes the payoff of a representative IEA member, P_s. At least each signatory is engaged in abatements $a_s > 0$, which typically exceed the global average $\overline{a}$. Therefore, we have the following first-order condition, where the marginal benefit is balanced against marginal costs and the impact on cost dispersion

$$S\left[\frac{1}{Sa_s + (N-S)\,a_o} - \frac{2\theta\,(a_s - \overline{a})}{N}\right] - 1 = 0 \tag{2}$$

Singletons, in contrast, behave non-cooperatively towards the coalition and the other outsiders. To determine their best responses, they maximize their own payoff P_j given the abatement strategies of all countries $i \neq j$. In equilibrium, symmetry implies that the abatement strategies of each outsider a_o will be the same. We obtain the first-order condition for the non-signatories

$$\frac{1}{Sa_s + (N-S)\,a_o} - \frac{2\theta\,(a_o - \overline{a})}{N} - 1 \leq 0 \tag{3}$$

While marginal costs both inside and outside the IEA are identical, marginal benefits between signatories and outsiders deviate by the factor S. Thus, due to the internalization of the environmental externality through the formation of an IEA, the abatement activity of a non-signatory falls short of that of a signatory state.

In equilibrium the abatement activities of the countries inside and outside the IEA are

$$a_s^*(S,\theta) = \begin{cases} \frac{1}{N-S+1} + \frac{1-S}{2\theta} + \frac{N(S-1)}{2\theta S} & \theta > \widetilde{\theta} \\ \text{for} & \\ \frac{\sqrt{N^4+8\theta N^2 S(N-S)}-N^2}{4\theta S(N-S)} & \theta \leq \widetilde{\theta} \end{cases} \tag{4}$$

and

$$a_o^*(S,\theta) = \begin{cases} \frac{1}{N-S+1} + \frac{1-S}{2\theta} & \theta > \widetilde{\theta} \\ \text{for} & \\ 0 & \theta \leq \widetilde{\theta} \end{cases} \tag{5}$$

The threshold level $\widetilde{\theta} = 0.5\,(S-1)\,(N-S+1)$ separates partial from complete free-riding of the outsiders when at least some signatories form an IEA.[21] While for sufficiently strong fairness

[20] An exception from uniqueness is given for $\theta = 0$. In that case, only aggregate abatements of the non-signatory states are unique, but we have a continuum of equilibria because of quasi-linearity. To simplify the analysis, in case of $\theta = 0$ we stick only to the symmetric solution.

[21] Note, even for countries with identical preferences, in equilibrium we end up with different participation strategies and thus asymmetric abatement levels (Equations (4) and (5)). Thus, although all countries are homogeneous *ex ante*, they become heterogeneous *ex post*.

considerations, $\theta > \widetilde{\theta}$, the non-signatories are partial free-riders, they behave as complete free-riders for rather weak fairness preferences, $\theta \leq \widetilde{\theta}$. Whether a country outside the IEA becomes a complete or partial free-rider depends on countries' preferences and the number of signatories.[22]

In equilibrium, for a given coalition size S, the aggregate abatement activity A corresponds to the sum of abatements of the signatories and outsiders

$$A\left(S,\theta\right) = Sa_s^*(S,\theta) + \left(N-S\right)a_o^*(S,\theta) \tag{6}$$

Subsequently, these equilibrium values are used to analyze the participation strategies at the first stage.

If preferences are purely selfish, an outsider is a complete free-rider. This result changes if justice and fairness enters the scene. In our framework, countries behave less selfish and more conforming through the choice of similar abatement strategies.

Proposition 1 *Abatements inside and outside the coalition.*
i) For signatories, stronger fairness preferences result in smaller abatement activities. If θ exceeds the threshold level $\widetilde{\theta}$ even an outsider becomes active. The stronger θ, the more abatements an outsider carries out. In the limit (for $\theta \to \infty$) there is no difference between an insider and an outsider.
ii) The aggregate does not significantly change in θ. For $\theta < \widetilde{\theta}$, fairness has a negative impact on global abatements, while $A(S,\theta)$ remains constant for all θ exceeding the threshold $\widetilde{\theta}$.[23]

Stronger fairness attitudes result in more homogeneity as countries inside and outside an IEA adjust their abatement levels to each other. In case of $\theta < \widetilde{\theta}$, outsiders are complete free-riders and more homogeneity requires a reduction in coalition's abatements. The price for less unequal cost sharing among countries is a loss in environmental quality as the global abatement measures $A\left(S,\theta\right)$ are reduced.

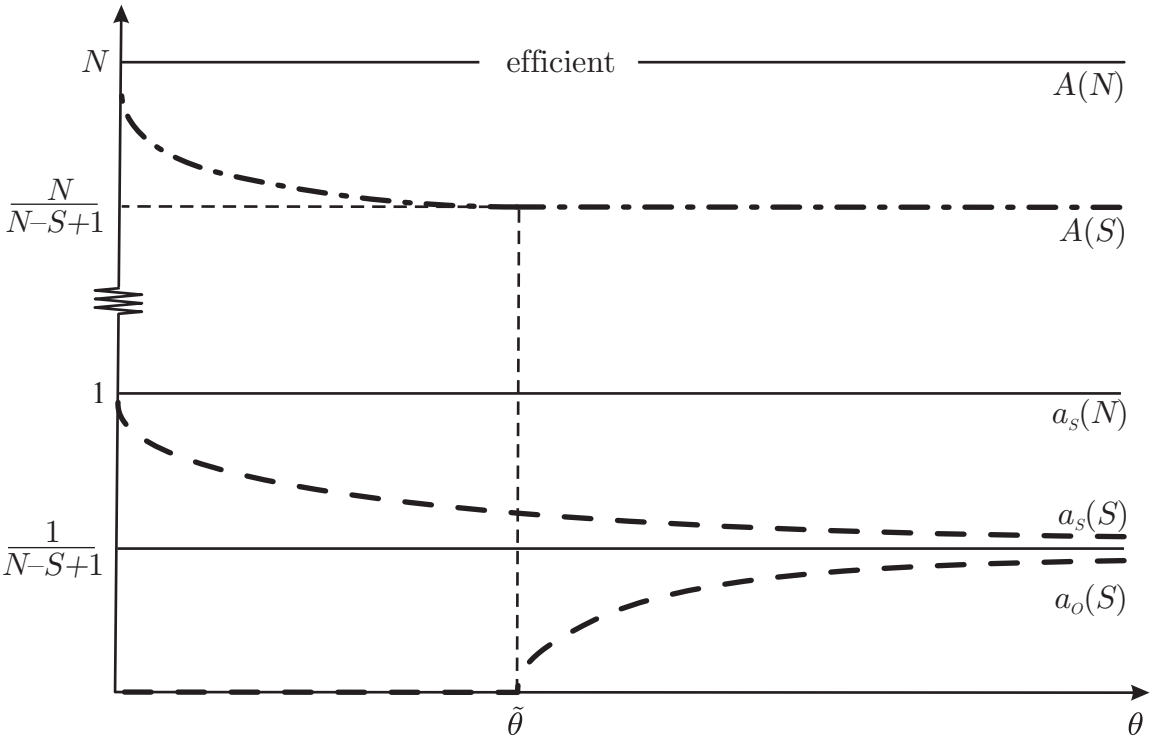
As long as both signatories and outsiders adopt an active measure, *i.e.,* for $\theta > \widetilde{\theta}$, the aggregate abatement $A(S,\theta)$ does not change. For stronger θ, countries' abatements become similar which results in a redistribution of cost shares from the members of an IEA to outsider countries. While the non-signatory states reinforce their environmental policy $a_o(S,\theta)$, measures of the IEA members $a_s(S,\theta)$ are reduced, see Figure 1. This behavior is driven by the wish not to deviate too much from the environmental policy the other countries realize. Thus, although countries choose different participation strategies they behave rather conform with respect to the abatement policy itself. Consequently, justice and fairness reduce the incentive to leave a coalition.

Summarizing, for each given coalition size S, justice and fairness enforce similar abatement strategies even at the expense of a reduction in overall measures. Conforming behavior in environmental policy results at the second stage of the entire game. However, what are the consequences for the participation strategies? Does the driving force for similar policy measures stabilize larger coalitions as governments feel a stronger incentive to join an agreement? As we are interested in the impact of our extended preferences on the size of an IEA, these questions will be analyzed in what follows.

[22] This threshold level increases with the total number of countries, $\partial\widetilde{\theta}/\partial N > 0$ respectively. The more countries that are faced with the environmental problem, the stronger is the incentive to free-ride. Thus, partial free-riding of the outsider countries requires a relatively strong θ. If the majority of the countries behave non-cooperatively (cooperatively), an increasing number of coalition members results in an increasing (decreasing) threshold level, $\partial\widetilde{\theta}/\partial S = \frac{N+1}{2} - S$.

[23] The proof for i) and ii) follows immediately from Equations (4), (5) and (6).

Figure 1. Abatements for coalition size S and N.



4. Signing an IEA: The First Stage

Introducing a measure for countries' fairness preferences means that similar behavior becomes decisive. Intuitively speaking, either nearly all, or almost none, of the countries are expected to sign an IEA. Whether this conjecture holds true will be analyzed in what follows.

Although the stability of an agreement depends on the payoffs resulting from the policy game at the second stage, it is worthwhile taking a closer look at the variance in the environmental policy, as this is the origin of our new insights into coalition formation.

4.1. Cost Dispersion

It is the interplay of S and θ which becomes decisive for cost dispersion measured by the variance in abatements. Both parameters have an impact on the extent of global abatements and the distribution of the cost shares as they determine countries' (partial or complete) free-riding behavior.

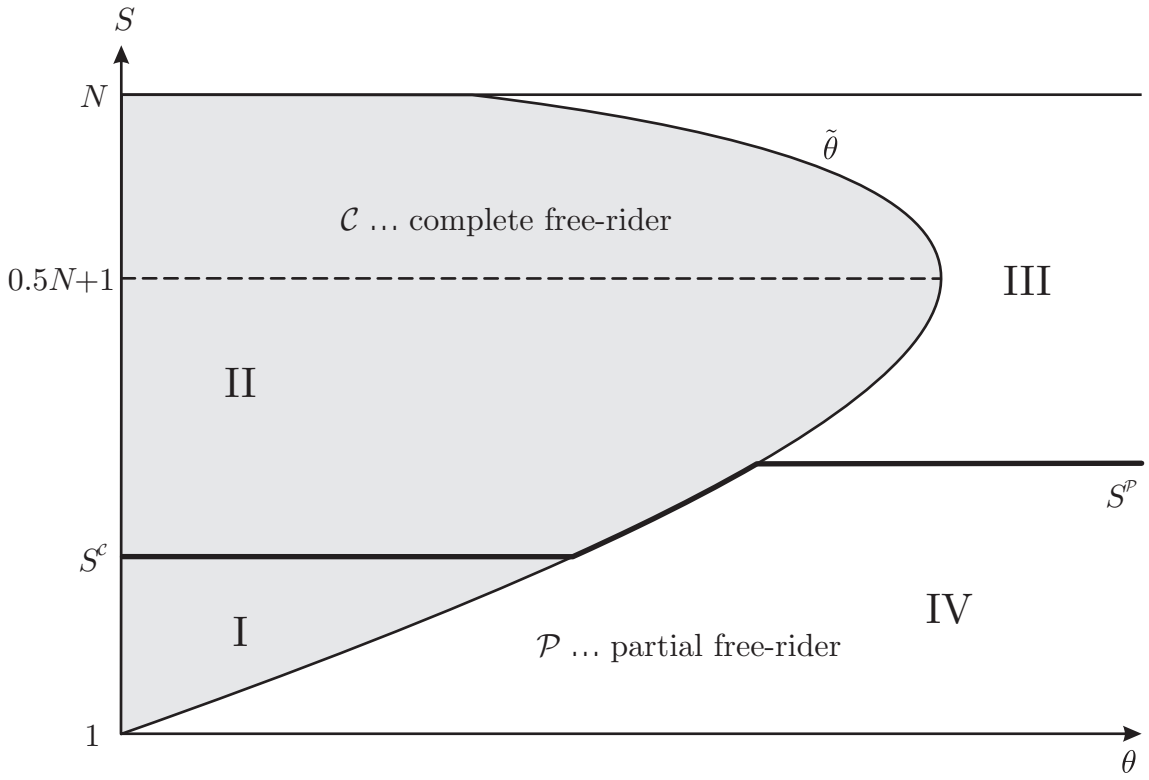
Inserting all information about the abatements (Equations 4 and 5), the variance is given by

$$\sigma\left(S,\theta\right)=\frac{\left(N-S\right)S\left(a_s-a_o\right)^2}{N^2}=\begin{cases}\frac{(N-S)(S-1)^2}{4\theta^2 S} & \theta>\widetilde{\theta}\\ \text{for} & \\ \frac{(N-S)S[a_s^*(S,\theta)]^2}{N^2} & \theta\leq\widetilde{\theta},\end{cases} \tag{7}$$

where the threshold level $\widetilde{\theta}$ is defined as before. Obviously, $\sigma(S,\theta)$ is continuous in both arguments and decreasing in θ as can be checked by a close look at Equations (4) and (7). Needless to say, cost dispersion vanishes for homogeneous behavior of all countries, *i.e.,* for identical participation strategies

at stage one, either $S = 1$ or $S = N$. Consequently, cost dispersion is non-monotonic in S as it changes with the relative size of the group of IEA members and non-signatories.

Figure 2. Free-riding and cost dispersion.



In the following, we prove the single-peakedness of σ in S. As there are two types of free-riding, we have to distinguish whether (S, θ) lies within the set characterizing complete free-riding $\mathcal{C} := \{(S, \theta) \mid \theta \leq \widetilde{\theta}\}$ or partial free-riding $\mathcal{P} := \{(S, \theta) \mid \theta > \widetilde{\theta}\}$. Figure 2 shows these two areas. In case of $(S, \theta) \in \mathcal{P}$ the variance $\sigma(S, \theta)$ is decreasing (increasing) in S if S exceeds a certain threshold $S^{\mathcal{P}} = \frac{N}{4} + \frac{N}{4}\sqrt{1 + \frac{8}{N}}$. This can be shown by maximizing σ with respect to S. According to (7), in case of partial free-riding the first-order condition

$$\frac{\partial \sigma}{\partial S} = \frac{S - 1}{4\theta^2 S^2}\left[NS - 2S^2 + N\right] = 0 \tag{8}$$

characterizes a unique solution for the maximum variance at $S^{\mathcal{P}}$.

For $(S, \theta) \in \mathcal{C}$ a similar property holds for a threshold $S^{\mathcal{C}} = 0.5N$, which can be seen if we rearrange the first-order condition in the case of complete free-riding by inserting (4)

$$[N - 2S]\frac{a_s^*}{\sqrt{N^4 + 8\theta N^2 S\,(N - S)}} = 0 \tag{9}$$

Obviously, we have $S^{\mathcal{C}} < S^{\mathcal{P}} < 0.5N + 1$, where the latter value characterizes the vertex of the curve $\widetilde{\theta}$ which separates the set $\mathcal{C}$ from $\mathcal{P}$ as shown in Figure 2.

We are now able to summarize our findings concerning cost dispersion. In areas I and IV $\sigma\,(S, \theta)$ is increasing in S, while it decreases in areas II and III. Thus, $\sigma\,(S, \theta)$ obtains its maximum $\sigma\,(S, \theta)$ at the bold line in Figure 2. For S below this line cost dispersion still increases in S, thereafter

it diminishes. Hence, for each given θ the variance $\sigma(S,\theta)$ is single-peaked in S. The following proposition summarizes our findings.

Proposition 2 *The impact of fairness and the size of an IEA on cost dispersion.*
The variance $\sigma(S,\theta)$ is decreasing in θ and single-peaked in S.

Obviously, the more the countries dislike an unfair cost dispersion, the smaller is the variance in their abatement activities. As fairness preferences aim at a conforming behavior of IEA members and non-signatories, they consequently reduce disparities in countries' cost shares for improving a global environmental problem.

What about the impact of IEA size on cost sharing? Intuitively speaking, starting from a small IEA, an increase in the coalition size yields more heterogeneity in countries' abatements as the variance accounts for all pair-wise differences in countries' policy measures, *i.e.,* the number of *insider meets outsider* increases until both groups are of nearly equal size. For $S \approx 0.5N$, we have two groups of nearly equal size, such that, with a further increase of S, the cost dispersion declines until it vanishes for $S = N$. If there is an overwhelming majority pro or con an IEA almost all countries adopt similar environmental measures, and thus, cost shares become rather similar too. For justice and fairness alone, a coalition consists ideally either of all countries or none. Thus, fairness and justice have a destabilizing impact on medium-sized coalitions.[24] Subsequently, we show how this effect changes the results on coalition formation previously analyzed in the traditional literature.

4.2. Stability Analysis

The stability analysis for coalition formation depends on the equilibrium payoff inside and outside the coalition that results from the policy game at stage 2, given the coalition size S. As all countries are assumed to be *ex ante* identical, each player's payoff depends on the coalition size and varies between signatories $P_s(S)$ and outsiders $P_o(S)$. The stability analysis is based on the status quo relative to the prevailing alternatives. External stability requires that no outsider has an incentive to join the coalition, *i.e.,* $P_s(S+1) - P_o(S) \leq 0$, while internal stability is fulfilled if no insider wants to leave the coalition, *i.e.,* $P_s(S) - P_o(S-1) \geq 0$. An extreme coalition formation with either $S = 1$ (complete failure of an IEA) or $S = N$ (grand coalition) only requires one of the above relations. For $S = 1$ it is sufficient to check external stability while for $S = N$ only internal stability matters.

For the equilibrium size S^* external and internal stability must be fulfilled simultaneously. These two payoff differences are implicit functions of θ and S. They can be solved analytically for θ which is described by a δ-function for $(S,\theta) \in \mathcal{P}$. Given δ the following equilibrium condition for partial free-riding can be stated

$$\delta(S^* - 1) \leq \theta \leq \delta(S^*) \tag{10}$$

The relation on the left side guarantees that no signatory wants to withdraw from the IEA, and the relation on the right prevents an outsider from joining the contract. If countries' offsetting behavior

[24] This result is in contrast to Hoel and Schneider [29], who extended governments preferences so that becoming an IEA member is an end in itself and encourages participation independent of coalition size under pure selfishness.

is characterized by partial free-riding, we obtain the following expression analytically by inserting (4) and (5) for $\theta > \widetilde{\theta}$:

$$\delta(S) = \frac{(N-1)[3S+1]S - 2S^3 - N}{4S(S+1)\left[\ln\frac{N-S+1}{N-S} - \frac{1}{(N-S)(N-S+1)}\right]} > 0 \tag{11}$$

Both the numerator and the denominator are strictly positive and finite for all $1 \leq S \leq N-1$. As partial free-riding is a prerequisite for the δ-function, $\delta(S) > \widetilde{\theta}(S)$ guarantees compatibility. In what follows, we will show that for a small number of countries affected by an international environmental problem, only partial free-riding is important and exclusively $\delta(S)$ proves for stability.

Lemma 3 *Local problems deal with partial free-riding.*
The stability of a coalition for $N < 12$ countries is exclusively determined by partial free-riding, while outsiders switch from partial to complete free-riding and back for $N \geq 12$.

Proof: The stability of a coalition is exclusively determined by partial free-riding if $\delta(S) > \widetilde{\theta}(S)$, while complete free-riding determines stability if the opposite holds. i) Inserting all integer numbers for $N \in [4, 12)$ shows that the relation $\delta(S) > \widetilde{\theta}(S)$ is satisfied for all S. ii) The relations $\delta(1) > \widetilde{\theta}(1)$ and $\delta(N-1) > \widetilde{\theta}(N-1)$ show that partial free-riding is always relevant for 1 and $N-1$, while $\delta(N-2) < \widetilde{\theta}(N-2)$ proves the relevancy of complete free-riding for $N-2$. Consequently, outsiders' behavior switches at least once from partial to complete free-riding and back. All three relations hold true for $N \geq 12$. Q.E.D.

According to Lemma 3 we distinguish between transboundary environmental problems, like the water quality in the Baltic sea, and global environmental problems, e.g., greenhouse gases. The number of involved countries in transboundary environmental problems is rather small, while global environmental problems affect many if not all countries. The externality increases in the number of involved countries and has an impact on the free-riding behavior of non-signatories. We have shown that for a relatively small number of countries concerned with an international problem complete free-riding does not occur, while both complete and partial free-riding play an important role for global environmental problems. Subsequently, we first analyze the stability for transboundary environmental problems and thereafter for global environmental problems.

Stability Analysis for Transboundary Environmental Problems

For $N < 12$ the environmental problem is rather local and bounded. In that case all countries are engaged in environmental measures, independent whether they have signed an IEA or not. Insider and outsider share the burden of environmental policy. Hence, cost dispersion is rather low.

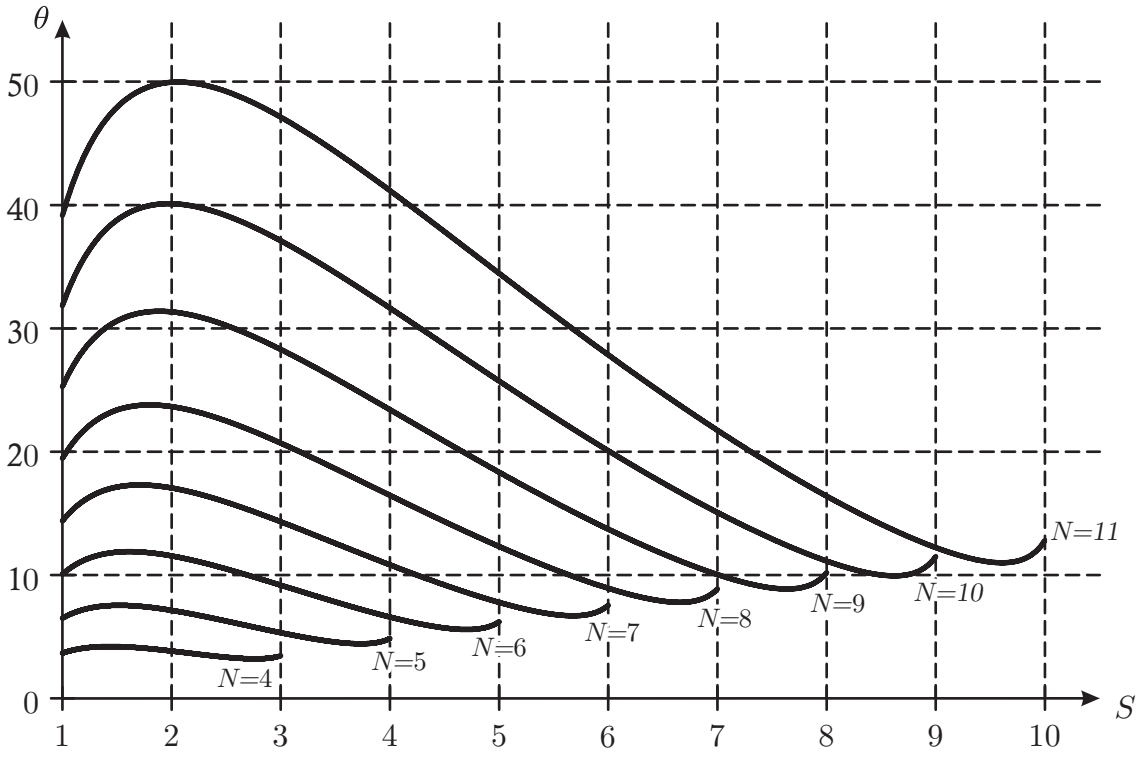
Given the δ-function, we can state the following equilibrium conditions for a small number of involved countries

$$\begin{array}{ll} \text{i) } \delta(S^*-1) \leq \theta \leq \delta(S^*) & \text{for interior equilibria } (2 \leq S^* \leq N-1), \\ \text{ii) } \theta \leq \delta(1) & \text{for failure of an IEA } (S^*=1), \text{ and} \\ \text{iii) } \delta(N) \leq \theta & \text{for the grand coalition } (S^*=N) \end{array} \tag{12}$$

In i) the left relation ensures that no signatory wants to withdraw from the IEA, and the right relation prevents an outsider from joining the contract. Obviously, except for corner solutions $S^* = 1$ or $S^* = N$, an interior equilibrium can only be stable where δ is increasing in S.

Transboundary environmental problems affect only few countries and therefore, coalitions are rather small. Preferences for justice and fairness aim at conforming behavior of both members and outsiders, and consequently, reduce disparities in countries' cost shares. Hence, each outsider behaves as partial free-rider and $\delta(S)$ proves for stability. Subsequently, we present our findings by drawing the δ-functions for $N \in [4, 12)$ in a diagram. As condition (12) has to be satisfied in equilibrium, a close look at Figure 3 shows all potential equilibria.

Figure 3. Stability analysis for $N \in [4, 12)$.



The δ-function is increasing at the beginning, *i.e.,* $\delta(2)$ always exceeds $\delta(1)$, while δ decreases thereafter for $S \geq 2$. Consequently, there doesn't exist any stable coalition between 2 and $N-1$. This result is an implication of cost dispersion's single-peakedness which destabilizes medium-sized coalitions. Furthermore, at its tail the δ-function is increasing. However, as coalition sizes can only be integer numbers we have to test for an all-but-one coalition whether $\delta(N-1)$ exceeds $\delta(N-2)$ or not. For $N \in [4, 8]$ there is no $S^* = N-1$ equilibrium as $\delta(N-2)$ exceeds $\delta(N-1)$, while the opposite holds for and $N \in [9, 12)$.[25]

As stable interior equilibria are located where the δ-function is increasing, we distinguish different areas for θ that are separated by $\delta(1)$, $\delta(2)$ and $\delta(N-1)$. Summarizing, we end up with four types of equilibria: two corner solutions ($S^* = 1$ and $S^* = N$) and two interior equilibria (the small coalition

[25]The numerical solution for $\delta(1), \delta(2), \delta(N-2)$ and $\delta(N-1)$ of the δ-function for $N \in [4, 12)$ is presented in the appendix.

$S^* = 2$ and an all-but-one coalition $S^* = N - 1$). The relevant thresholds for θ that distinguish these areas for a failure of an IEA and the small coalition are $\delta(1)$ and $\delta(2)$. The stability of the grand coalition $S^* = N$ requires $\theta > \delta(N-1)$. Our findings are summarized in the next proposition.

Proposition 4a *Stable equilibria for transboundary problems,* i.e., $N \in [4, 12)$.
i) A complete failure of an IEA $(S^* = 1)$ *is stable for* $\theta \in [0, \delta(1)]$.
ii) The stability of the grand coalition $S^* = N$ *requires* $\theta > \delta(N-1)$.
iii) For $\theta \in (\delta(1), \delta(2))$ *we obtain a small coalition* $S^* = 2$.
iv) The all-but-one coalition $S^* = N - 1$ *becomes stable* $\theta \leq \delta(N-1)$ *and* $N \in [9, 12)$. *For* $N \in [4, 8]$ *this type of stable IEA does not exist.*

Proof: All results directly follow from the relation $\min\{\delta(N-2), \delta(N-1)\} < \delta(1) < \delta(2)$ and table A in the appendix. Q.E.D.

Preferences for justice and fairness aim at conforming behavior of both members and outsiders. Consequently, all countries reduce disparities in cost shares. As proposition 4a shows, a stable coalition in case of a local environmental problem, *i.e.,* $N \in [4, 12)$, can never be medium-sized. Justice and fairness stabilize not only the small, but even larger coalitions. All kinds of equilibria are rather like those in the *battle of the sexes.* Adopting your partner's participation strategy is better than any other behavior.

Stability Analysis for Global Environmental Problems

As previously presented for a small number of countries affected by an international problem partial free-riding is decisive for outsider's behavior. However, as the externality increases with the number N of involved countries the incentive for outsiders not to provide any abatements becomes stronger. Thus, according to Lemma 3, both complete and partial free-riding are relevant for global environmental problems $(N \geq 12)$. This result shows that for global environmental problems both types of free-riding must be dealt with endogenously.[26]

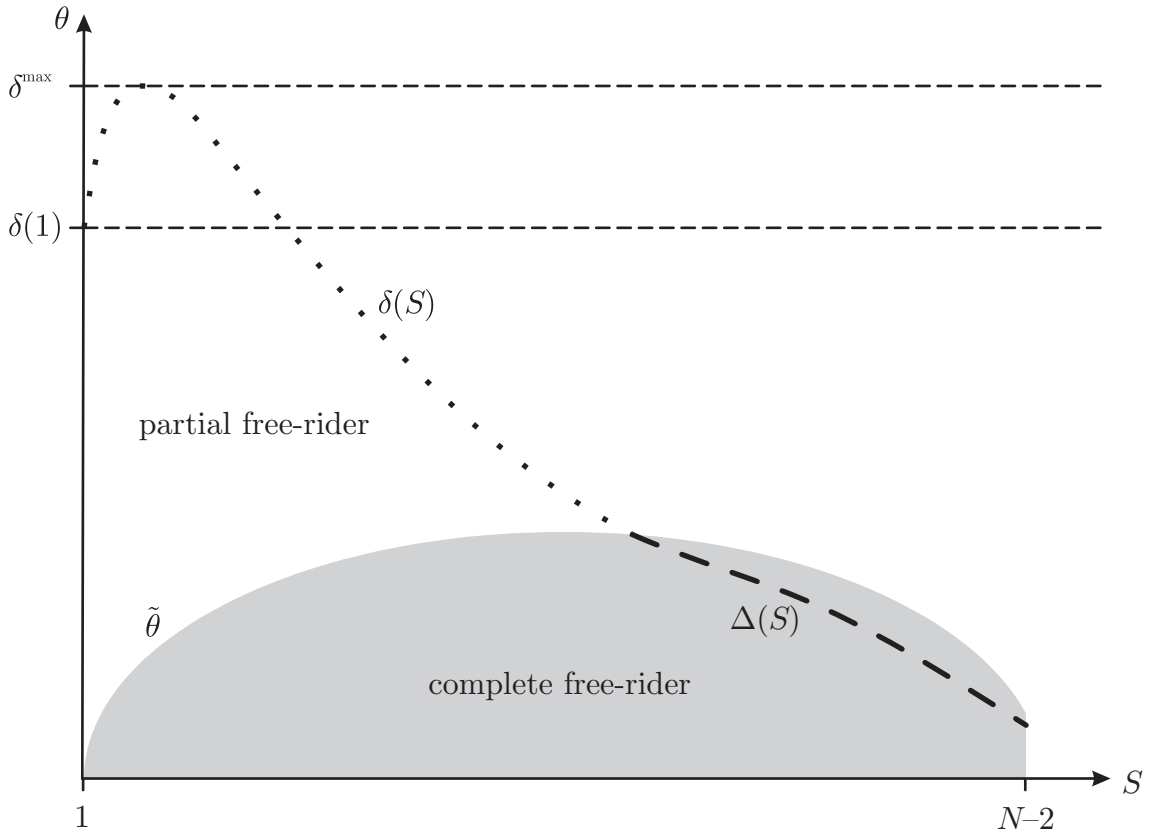
As mentioned before, the offsetting behavior of the potential outsider in the case of larger coalitions (grand $S = N$ or all-but-one $S = N - 1$) switches from partial to complete free-riding. Thus, starting from the grand coalition, a single outsider reduces its abatements but remains active, while, with two outsiders, both become complete free-riders. Furthermore, for a rather small IEA outsider behave again as partial free-riders. Thus, if almost all countries or almost none sign an IEA the homogeneity yields a low cost dispersion. Starting from a small IEA, an increase in the coalition size implies more heterogeneity in countries' abatements until both groups are of nearly equal size and declines until it vanishes for the grand coalition. For medium-sized coalitions, when countries adopt rather heterogeneous decisions with respect to signing an IEA, cost shares are rather heterogeneous, too, and outsiders become complete free-riders. It is the interplay of S and θ that has an impact on the extent of global abatements and the distribution of the cost shares. Subsequently, we show which impact the switch in the free-riding behavior has on coalition formation for global environmental problems.

[26] This complements Lange and Vogt [31], who neglected the endogeneity of these types of free-riding.

As we consider two different kinds of free-riding, we have to analyze the participation conditions for both types of offsetting behavior outside the IEA. Similar to the δ-function for partial free-riding, we deduce a Δ-function for complete free-riding. However, there doesn't exist an analytical solution for Δ. Thus, we have to stick to a numerical solution in case of complete free-riding.[27] To consider the offsetting behavior outside the IEA appropriately, the equilibrium condition for $N \geq 12$ is determined by the combined δ-Δ-function. Given this function as drawn in Figure 4, we can prove for the stability of an IEA according to similar conditions as in (12). Thus, except for corner solutions $S = 1$ or $S = N$, an interior equilibrium can only be stable where δ respectively Δ is increasing in S.

Subsequently, we present our findings by using a numerical example with $N = 100$. We have checked the shape of the δ-Δ-function for all integer numbers N from 12 up to 200 and found that all our results remain qualitatively unchanged.

Figure 4. Stability analysis for $N = 100$.



The following analysis contains two parts. Figure 4 presents the δ-Δ-function for coalition sizes up to $(N-2)$. It shows the relevant parts of the functions $\delta(S)$ as a dotted line and $\Delta(S)$ as a dashed line that are separated from each other through $\widetilde{\theta}$. While the δ-part shows a maximum, the Δ-function is monotonously decreasing up to $(N-2)$. Within this range, offsetting behavior changes from partial to complete free-riding.

[27] The implicit Δ-function and its numerical solution are presented in the appendix.

Proposition 5 *There are no stable equilibria with outsiders that behave as complete free-riders.*

Proof: As the Δ-function is strictly decreasing in S, there is no scope for stable interior equilibria when outsiders are complete free-rider. Q.E.D.

As soon as outsiders are rather selfish and behave as complete free-riders, the distribution of the abatement costs is relatively heterogeneous among countries. This will be regarded as unfair cost sharing of the abatement costs among the involved countries and prevents the formation of stable equilibria.

Furthermore, according to the proof of Lemma 3, offsetting changes back to partial free-riding when almost all countries join the IEA, *i.e.,* from $(N-2)$ to $(N-1)$. Here, we have to distinguish two cases: According to our simulations, for $N < 28$ the δ-Δ-function increases in the relevant range, while it declines for $N \geq 28$.

As stable interior equilibria are located where the combined δ-Δ-function is increasing, we distinguish different areas for θ that are separated by $\delta(1)$, $\delta^{\max}$, $\Delta(N-2)$, and $\delta(N-1)$, where $\delta^{\max}$ is defined as the maximum of $\delta(S)$ for integer $S < N$. Similar to transboundary environmental problems, we end up with four types of equilibria: two corner solutions ($S^* = 1$ and $S^* = N$) and two interior equilibria (a small coalition and the all-but-one coalition $S^* = N-1$). According to our numerical solution small coalitions for $N \leq 200$ do not exceed 9 signatories. For the all-but-one coalition, we have to check whether $\delta(N-1)$ exceeds $\Delta(N-2)$. In either case, $\delta(1)$ and $\delta^{\max}$ are the relevant thresholds for θ that distinguish the areas for a failure of an IEA and a small coalition.

There are two different cases: i) If $\Delta(N-2)$ falls short of $\delta(N-1)$ there is—as in Figure 5—scope for all-but-one equilibria as well as for the grand coalition. ii) The opposite holds in Figure 6. Thus, we can summarize our findings in the next proposition.

Proposition 4b *Stable equilibria for global environmental problems, i.e.,* $N \geq 12$.
i) The corner solution $S^* = 1$ *is stable for* $\theta \in [0, \delta(1)]$.
ii) The stability of the grand coalition $S^* = N$ *requires* $\theta > \delta(N-1)$.
iii) For $\theta \in (\delta(1), \delta^{\max})$ *we obtain a rather small coalition* $S^* > 1$.
iv) The all-but-one coalition $S^* = N-1$ *is stable if* $\delta(N-1)$ *exceeds* $\Delta(N-2)$
and $\theta \in [\Delta(N-2), \delta(N-1)]$. *The interval is non-empty for* $N < 28$. *For* $N \geq 28$ *this type of stable IEA does not exist.*

Proof: As the relation $\min\{\Delta(N-2), \delta(N-1)\} < \delta(1) < \delta^{\max}$ holds true in any case, we can distinguish the four areas for θ given in (i) up to (iv). The first relation follows from inserting all information in the δ function, the second from the property of a maximum. The non-empty interval for $N < 28$ in (iv) can only be checked numerically, which we did for $N \leq 200$. Q.E.D.

Figure 5. Equilibria with all-but-one for $12 \leq N < 28$.

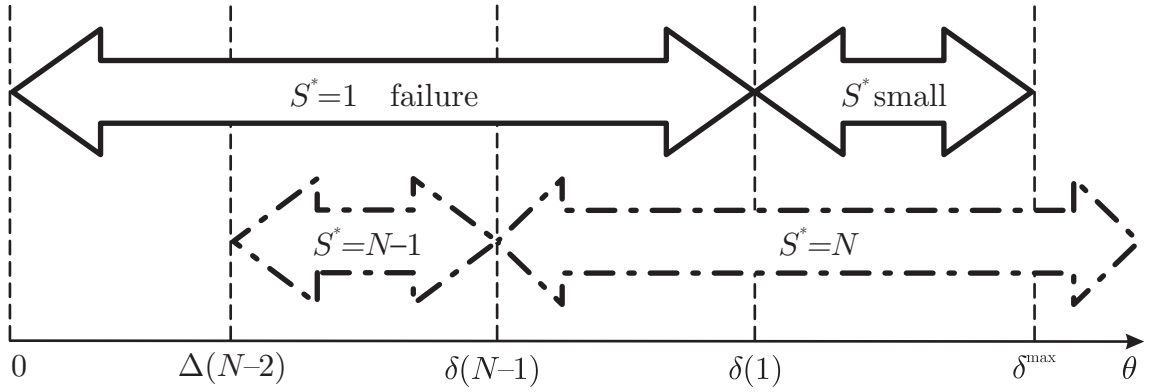
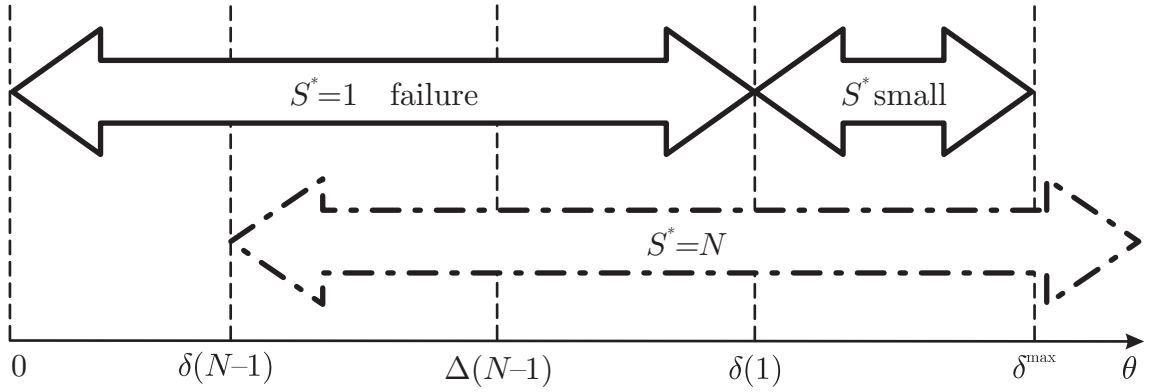


Figure 6. Equilibria without all-but-one for $N \geq 28$.



As proposition 4a and b show, a stable coalition with preferences for justice and fairness can never be medium-sized. A unique equilibrium of the entire game only exists for rather low and very high θ. The subgame-perfect equilibrium is unique for the grand coalition if $\theta > \delta^{\max}$ or the complete failure of an IEA if $\theta < \min\{\Delta(N-2), \delta(N-1)\}$. If θ is somehow in between, coalitions always coexist where either almost all countries and almost none participate in an IEA.

In general, there are three effects which determine stability: *individual free-riding*, *collective internalization*, and *conforming behavior*, where the first two are Barrett's [4,5] traditional effects. The individual gain from free-riding is rather selfish and depends on the deviation between insider's and outsider's abatement costs. The gain from internalization can be measured by the change in global abatements when a single country leaves or enters the coalition. Finally, justice and fairness favor similar policy measures, irrespective of the chosen participation strategy. According to the single-peakedness of the variance in S, conforming behavior destabilizes medium-sized coalitions as they provide an incentive either to leave or to join a coalition. Free-riding favors leaving a coalition, the efficiency argument goes in the opposite direction, while fairness prefers homogeneous behavior. Consequently, it is the interplay of these three effects that determines the equilibrium coalition size.

Justice and fairness favor coalitions where either almost all or almost none sign an IEA. While for small coalitions conforming behavior and free-riding are complementary in providing an incentive for staying outside an IEA, when the coalition size becomes large they favor opposite participation

strategies. As in the traditional literature, it is the internalization gain which stabilizes IEAs, while the free-riding effect hinders larger coalitions. Thus, stronger fairness preferences are needed to overcome the instability of the grand coalition. In case of an all-but-one coalition, the free-riding gain works to destabilize the grand coalition, but it is not strong enough to countervail the ones from internalization and conforming behavior.[28]

5. Extension[29]

The previous equilibrium analysis of the basic model focused on identical countries and symmetric preferences. However, in real life countries differ with respect to their cost and benefit structure or fairness considerations. Therefore, the remainder of this paper allows for some asymmetries to generalize our results. We first analyze the impact of asymmetric equity preferences and subsequently allow for more heterogeneity among countries.

5.1. Asymmetries in Measuring Equity

In what follows, we study the implications of different disutilities that result if the country itself or other countries abate more or less than the average. First, we analyze countries with a priority regarding their self-interest and second, advantageous and disadvantageous inequality. Note that both subsequently analyzed cases still assume identical countries.

Priority of Countries' Self-interest

Countries are motivated by issues of justice and fairness but still have self-interests. Therefore, we assume that countries regard their own deviations from the average different than deviations of all other countries. Then, the previous measure of justice and fairness σ is modified by integrating countries' self-interest. The heterogeneity in the abatement costs is measured by $\widetilde{\sigma}_1 = \sigma + b\left(a_j - \overline{a}\right)^2$, where $a_j \geq 0$ corresponds to the own abatement level of country j, and b measures the relative intensity of own deviations from the average. For $b > 0$ own deviations are regarded as more important than deviations of all other countries. Countries' behavior is driven by the wish not to deviate too much from the environmental policy of other countries. Thus, in equilibrium the abatement activities of the countries inside an IEA a_s^* decrease, while outsiders reinforce their abatement policy a_o^*. As a consequence, countries' self-interest enforces conforming behavior with respect to the abatement policy although countries chose different participation strategies. Moreover, cost shares are redistributed from the members of the IEA to the outsider countries. This reduces the cost dispersion and the incentive to leave a coalition. The opposite holds for $b < 0$.[30]

[28] In the case of complete free-riding a non-signatory's choice is restricted to $a_o = 0$. Hence, the gain from free-riding becomes smaller than for interior abatement strategies. As free-riding behavior changes from $(N-2)$ to $(N-1)$ this stabilizes the all-but-one coalition.

[29] We owe this section to a hint of the anonymous referee who suggested to abandon the strong assumption of symmetry in countries as well as in symmetry of inequality aversion.

[30] The case $b = 0$ corresponds to the previously analyzed model with symmetric equity preferences.

Even with a more intense self-interest cost dispersion remains single-peaked which, in turn, is the driving force behind our equilibrium results at stage one. Intuitively, an intensified conform behavior qualifies a coalition formation of almost all or nearly none.

Advantageous and Disadvantageous Inequality

According to the approaches of Fehr and Schmidt [16] and Bolton and Ockenfels [14] countries are assumed to suffer more from disadvantageous inequality than from advantageous inequality, *i.e.,* cost dispersion is non-symmetric to the average. A disadvantageous inequality holds true when some countries abate below the average, while the opposite characterizes an advantageous inequality. Typically, the latter is not as bad as the former. This is in line with empirical data from experimental economics on fairness preferences from Blanco *et al.* [35] and Dannberg *et al.* [36] who provide evidence that advantageous and disadvantageous inequality are seen differently. Through modifying the measure for justice and fairness σ by incorporating advantageous inequality, the heterogeneity in the abatement costs corresponds to $\widetilde{\sigma}_2 = \sigma + B \sum_{a_i<\overline{a}} \frac{1}{N} \left(a_i - \overline{a}\right)^2$. Then, $B > 0$ measures the intensity of the disadvantageous inequality relative to above average behavior. A country j regards the variance in all abatement strategies but additionally dislike the disadvantageous inequality, *i.e.,* all $a_i < \overline{a}$.[31] However, the impact on the abatement and participation strategies in equilibrium is similar to the former case.[32]

Both previously analyzed modifications assume identical countries with asymmetries in equity preferences. However, the analysis reveals that the abatement and participation strategies for asymmetric preferences remain qualitatively unchanged relative to our basic framework.

5.2. Heterogeneous Countries

In what follows, we give up the assumption of identical countries. We study two different scenarios, where countries differ either in their equity preferences or abatement costs.

Different Equity Preferences

To keep the analysis simple, we assume that countries' equity preferences can only be of two types, either high or low with $\theta^h > \theta^l$.[33] Subsequently, the subscript l and h distinguish low and high preference countries. The driving force of justice and fairness is each countries' wish not to deviate too much from the environmental policy of other countries. The more the countries dislike an unfair cost dispersion, the smaller is the own deviation from the global average. Thus, in equilibrium insider and outsider of both types show different abatement activities with $a_s^l > a_s^h > a_o^h \geq a_o^l$. Moreover, countries differ only in their preference intensity but the measure for justice and fairness σ is identical for all countries and remains single-peaked in the coalition size. Thus, two effects result. First, high preference countries feel a stronger incentive to adopt similar participation strategies. They favor the decision of the majority pro or con an IEA, where either almost all or nearly none are signatories. Then, analogous to the case of

[31] In contrast to the former assumption the deviation of a specific group and not only that of a single country counts differently.

[32] The impact of disadvantageous inequality for $B < 0$ can be modeled similarly.

[33] See Lange and Vogt [31] for an similar analysis of the extreme case in which countries are either interested in their absolute payoff ($\theta = 0$) or exclusively in equity ($\theta \to \infty$).

symmetric countries, we end up with rather small or large coalitions, while medium size coalitions are destabilized. Second, high preference countries feel a stronger incentive to join an IEA as they suffer more from a variance in all abatement strategies. Intuitively speaking, if both groups are of equal size the majority of the IEA members have high preferences. As both effects overlap, it is their interplay which determines the coalition size in equilibrium. However, although different equity preferences modify the results, the single-peakedness of the variance remains the driving force behind coalition formation.

Different Abatement Costs

The following model assumes that the countries involved differ in their cost structure. We distinguish countries with low and high marginal abatement costs, *i.e.,* $c^h \geq c^l > 0$. However, when countries differ in their abatement costs they also have a different perception of what seems to be 'fair'. Fairness puts some pressure on governments to accept similar responsibilities, especially for countries of the same type. Therefore, we assume that justice and fairness preferences focus on reducing the cost dispersion within countries of the same type by measuring the variance in all abatement strategies within this group. Thus, fairness measures whether equals are treated equally. In that case, a low and high cost country's payoff (P_j^l, P_j^h) is given as benefit minus costs minus a term which measures heterogeneity in all abatement strategies of the own group

$$P_j^l = \ln\left(\sum\nolimits_i a_i\right) - c^l \cdot a_j - \theta \cdot \widetilde{\sigma}^l \quad \text{and} \quad P_j^h = \ln\left(\sum\nolimits_i a_i\right) - c^h \cdot a_j - \theta \cdot \widetilde{\sigma}^h \tag{13}$$

Cost dispersion within the set of low cost countries L is measured by $\widetilde{\sigma}^l = \sum_{i \in L} \frac{1}{L} \left(a_i - \overline{a}_l\right)^2$, where $\overline{a}_l$ is the average of all low cost countries' environmental policies. Analogous, we define $\widetilde{\sigma}^h$ for the set of high cost countries H.

Low cost countries have a decisive cost advantage. Therefore, they feel a stronger incentive to join an IEA, while high cost countries have a stronger incentive to leave an IEA.[34] As a consequence, low cost countries with their cost advantage are easier to attract for an IEA than high cost countries. Moreover, due to equity preferences countries do not wish to deviate too much from the environmental policy of the other countries of the same type. Thus, countries of type (l, h) choose similar abatement and participation strategies within their group. Then, similar to our basic model, in equilibrium almost all or nearly none of each type join an IEA.

Summarizing, all extensions above modify our basic model in a continuous way through an additional economic parameter. Thus, the equilibrium outcome of the abatement game changes only slightly. Additionally, the single-peakedness of the variance in the number of IEA members stays unaltered and remains the driving force behind coalition formation. Consequently, our results can be expected to be robust against introducing some asymmetries and moderate heterogeneities.

6. Concluding Remarks

In the standard literature on International Environmental Agreements (IEA) empirical and theoretical predictions are inconsistent if we focus on the number of signing countries. While theory only proves

[34] See Buchholz *et al.* [39] for a stability analysis of self-enforcing IEAs for heterogeneous countries which differ in their abatement costs.

the existence of small coalitions, there is empirical evidence for larger agreements such as the Kyoto or Montreal Protocols. By extending countries' preferences to incorporate issues of fairness and justice, governments try to avoid welfare losses due to cost dispersion, measured by the variance in countries' abatement policies. Such preferences provide an incentive for countries to behave in the same way, either almost all, or almost none of the countries form an IEA. In both cases, the participation decisions are similar which stabilizes both, larger and smaller coalitions but destabilizes medium-sized coalitions.

In reality countries are not identical in every respect. Thus, in an extension we prove that our basic model is robust against an assumption of heterogeneous countries and more realistic preferences where equity concerns are asymmetric. A countries' cost share above the global average is much better than a similar deviation in the other direction.

Acknowledgements

Helpful comments by Michael Finus, Alexander Haupt, Silke Gottschalk, Michael Grüning and the participants of the IIPF and EAERE conference as well as workshops in Berlin, Bonn, and Rostock are gratefully acknowledged. We also would like to thank an anonymous referee for helpful comments and the German Research Foundation (DFG) for support through the SPP 1142 program on "Institutional Design of Federal Systems".

References

1. Murdoch, J.C.; Sandler, T. The voluntary provision of a pure public good: The case of reduced CFC emissions and the Montreal protocol. *J. Public Econ.* **1997**, *63*, 331–349.
2. Böhringer, C.; Vogt, C. Economic and environmental impacts of the Kyoto Protocol. *Can. J. Economics* **2003**, *36*, 475–494.
3. Böhringer, C.; Vogt, C. Dismantling of a breakthrough: The Kyoto Protocol–just symbolic policy. *Eur. J. Polit. Econ* **2004**, *20*, 597–617.
4. Barrett, S. International environmental agreements as games. In *Conflicts and Cooperation in Managing Environmental Resources*; Pethig, R., Ed.; Springer: Berlin, Germany, 1992; pp. 11–37.
5. Barrett, S. Self-enforcing international environmental agreements. *Oxford Econ. Pap.* **1994**, *46*, 878–894.
6. Carraro, C.; Siniscalco, D. Strategies for the international protection of the environment. *J. Public Econ.* **1993**, *52*, 309–328.
7. Finus, M. *Game Theory and International Environmental Cooperation*; Edward Elgar: Cheltenham, UK, 2001.
8. Hoel, M. International environmental conventions: The case of uniform reductions of emissions. *Environ. Resour. Econ.* **1992**, *2*, 141–159.
9. Barrett, S. Consensus treaties. *J. Inst. Theor. Econ.* **2002**, *158*, 529–554.
10. Barrett, S.; Stavins, R. Increasing participation and compliance in international climate change agreements. *Int. Environ. Agreem. - P.* **2003**, *3*, 349–376.
11. Buchholz, W.; Peters, W. A rawlsian approach to international cooperation. *Kyklos* **2005**, *58*, 25–44.

12. Stern, N. *The Economics of Climate Change: The Stern Review*; Cambridge University Press: Cambridge, UK, 2007.
13. Alesina, A.; Angeletos, G.M. Fairness and redistribution. *Am. Econ. Rev.* **2005**, *95*, 960–980.
14. Bolton, G.E.; Ockenfels, A. ERC: A theory of equity, reciprocity, and competition. *Am. Econ. Rev.* **2000**, *90*, 166–193.
15. Falk, A.; Fehr, E.; Fischbacher, U. Reasons for conflict: Lessons from bargaining experiments. *J. Inst. Theor. Econ.* **2003**, *159*, 171–187.
16. Fehr, E.; Schmidt, K. A theory of fairness, competition and cooperation. *Q. J. Econ.* **1999**, *114*, 817–68.
17. Rabin, M. Incorporating Fairness into game theory and economics. *Am. Econ. Rev.* **1993**, *85*, 1281–1302.
18. Sobel, J. Interdependent preferences and reciprocity. *J. Econ. Lit.* **2005**, *43*, 392–436.
19. Postlewaite, A. The social basis of interdependent preferences. *Eur. Econ. Rev.* **1998**, *42*, 779–800.
20. Ringius, L.; Torvanger, A.; Underdal, A. Burden sharing and fairness principles in international climate policy. *Int. Environ. Agreem. - P.* **2002**, *2*, 1–22.
21. Lange, A.; Vogt, C.; Ziegler, A. On the importance of equity in international climate policy: An empirical analysis. *Energ. Econ.* **2007**, *29*, 545–562.
22. Lange, A.; Löschel, A.; Vogt, C.; Ziegler, A. On the self-serving use of equity principles in international climate negotiations. *Eur. Econ. Rev.* **2010**, *54*, 359–375.
23. Albin, C. Negotiating international cooperation: global public goods and fairness. *Rev. Int. Stud.* **2003**, *29*, 365–385.
24. Lindbeck, A. Incentives and social norms in household behavior. *Am. Econ. Rev.* **1997**, *87*, 370–377.
25. Wooders, M.; Cartwright, E.; Selten, R. Behavioral conformity in games with many players. *Game. Econ. Behav.* **2007**, *57*, 347–360.
26. Elster, J. Social norms and economic theory. *J. Econ. Perspect.* **1989**, *3*, 99–117.
27. Rege, M. Social norms and private provision of public goods. *J. Public Econ. Theory* **2004**, *6*, 65–77.
28. Finus, M.; Rundshagen, B. Towards a positive theory of coalition formation and endogenous instrumental choice in global pollution control. *Public Choice* **1998**, *96*, 145–186.
29. Hoel, M.; Schneider, K. Incentives to participate in an international environmental agreement. *Environ. Resour. Econ.* **1997**, *9*, 153–170.
30. Jeppesen, T.; Andersen, P. Commitment and fairness in environmental games. In *Game Theory and the Environment*; Hanley, N., Folmer, H., Eds.; Edward Elgar: Cheltenham Northampton, Massachusetts, USA, 1998; pp. 65–83.
31. Lange, A.; Vogt, C. Cooperation in international environmental negotiations due to a preference for equity. *J. Public Econ.* **2003**, *87*, 2049–2067.
32. Victor, D.G.; Coben, L.A. A herd mentality in the design of international environmental agreements? *Global Environ. Polit.* **2005**, *5*, 24–57.
33. Rawls, J. *A Theory of Justice*; Revised Edition; Belknap Press of Harvard University Press: Cambridge, USA, 1999.

34. Engelmann, D.; Strobel, M. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *Am. Econ. Rev.* **2004**, *94*, 857–869.
35. Blanco, M.; Engelmann, D; Normann, H.-T. A within subject analysis of other-regarding preferences. Royal Holloway College, University of London, Working Paper, 2005.
36. Dannberg, A.; Riechmann, T.; Sturm, B.; Vogt, C. Inequity aversion and individual behavior in public good games: An experimental investigation. *ZEW discussion paper* **2007**, No. 07–034.
37. d'Aspremont, C.A.; Gabszewicz, J.J. On the stability of collusion. In *New Developments in the Analysis of Market Structure*; Stiglitz, J.E., Mathewson, G.F., Eds.; Macmillan: New York, NY, USA, 1986; pp. 243–264.
38. Finus, M.; v. Mouche, P; Rundshagen, B. Uniqueness of coalitional equilibria. *FEEM discussion paper* **2005**, No. 2005.23.
39. Buchholz, W.; Grüning, C.; Peters, W. Can heterogeneous countries form homogeneous international environmental agreements? European University Viadrina, Working Paper **2009**.

Appendix: δ-Δ-function

1. The numerical solution for $\delta(1), \delta(2), \delta(N-2)$ and $\delta(N-1)$ of the δ-function for $N \in [4, 12)$ are:

N	$\delta(1)$	$\delta(2)$	$\delta(N-2)$	$\delta(N-1)$
4	3.670	3.839	3.839	3.452
5	6.498	7.137	5.322	4.854
6	10.068	11.551	6.596	6.213
7	14.385	17.060	7.782	7.550
8	19.451	23.656	8.927	8.876
9	25.266	31.338	10.039	10.193
10	31.830	40.105	11.138	11.505
11	39.144	49.956	12.226	12.814

2. For complete free-riding $(S, \theta) \in \mathcal{C}$ we have to solve $P_S(S+1) - P_o(S) = 0$ by inserting the equilibrium activities (4) and (5) for $\theta \leq \widetilde{\theta}$. Signatories' measures $a_s(S, \theta)$ and $a_s(S+1, \theta)$ and consequently the payoff difference are implicit functions of θ and S. As the payoff differences can not be solved for θ explicitly a numerical solution will be used instead. This numerical solution is called Δ. A numerical solution of the above payoff difference is calculated for all integer $12 \leq N \leq 200$ and all applicable $1 < S < (N-2)$. As a result we obtain all θ where $P_S(S+1) - P_O(S) = 0$. However, not all θ lead to $(S, \theta) \in \mathcal{C}$. Therefore, in a second step all resulting θ will be tested against the complete free-riding condition.