Eichner, Thomas; Pethig, Rüdiger

**Working Paper**

# Self-enforcing environmental agreements and international trade

Volkswirtschaftliche Diskussionsbeiträge, No. 156-12

**Provided in Cooperation with:**
Fakultät III: Wirtschaftswissenschaften, Wirtschaftsinformatik und Wirtschaftsrecht, Universität Siegen

*Suggested Citation:* Eichner, Thomas; Pethig, Rüdiger (2012) : Self-enforcing environmental agreements and international trade, Volkswirtschaftliche Diskussionsbeiträge, No. 156-12, Universität Siegen, Fakultät III, Wirtschaftswissenschaften, Wirtschaftsinformatik und Wirtschaftsrecht, Siegen

This Version is available at:
https://hdl.handle.net/10419/84978

UNIVERSITÄT SIEGEN

Fakultät III
Wirtschaftswissenschaften,
Wirtschaftsinformatik und
Wirtschaftsrecht

Econ

Volkswirtschaftliche Diskussionsbeiträge

Discussion Papers in Economics

No. 156-12

September 2012

**Thomas Eichner · Rüdiger Pethig**

# Self-enforcing environmental agreements and international trade

Universität Siegen
Fakultät III
Wirtschaftswissenschaften, Wirtschaftsinformatik und Wirtschaftsrecht
Fachgebiet Volkswirtschaftslehre
Hölderlinstraße 3
D-57068 Siegen
Germany

http://www.uni-siegen.de/fb5/vwl/

ISSN 1869-0211

Available for free from the University of Siegen website at
http://www.uni-siegen.de/fb5/vwl/research/diskussionsbeitraege/

Discussion Papers in Economics of the University of Siegen are indexed in RePEc
and can be downloaded free of charge from the following website:
http://ideas.repec.org/s/sie/siegen.html

# Self-enforcing environmental agreements and international trade*

Thomas Eichner
Department of Economics, University of Hagen

Rüdiger Pethig
Department of Economics, University of Siegen

**Abstract**

In the *basic model* of the literature on international environmental agreements (IEAs) (Barrett 1994; Rubio and Ulph 2006) the number of signatories of self-enforcing IEAs does not exceed three, if non-positive emissions are ruled out. We extend that model by introducing a composite consumer good and fossil fuel that are produced and consumed in each country and traded on world markets. When signatory countries act as Stackelberg leader and emissions are positive, the size of stable IEAs may be significantly larger in our model with international trade. This would be good news if larger self-enforcing IEAs would lead to stronger reductions of total emissions. Unfortunately, the allocation of total emissions in self-enforcing IEAs turns out to be approximately the same as in the business as usual scenario independent of the number of its signatories. We also investigate the role of international trade by comparing our free-trade results with the outcome in the regime of autarky. Our autarky model turns out to coincide with the basic model of the literature alluded to above. We contribute to that literature by showing that in autarky regime the outcome of self-enforcing IEAs is also approximately the same as in business as usual.

JEL classification: C72, F02, Q50, Q58
Key words: international trade, self-enforcing environmental agreements
Stackelberg equilibrium

*Eichner: Department of Economics, University of Hagen, Universitätsstr. 41, 58097 Hagen, Germany, email: thomas.eichner@fernuni-hagen.de; Pethig: Department of Economics, University of Siegen, Hölderlinstr. 3, 57068 Siegen, Germany, pethig@vwl.wiwi.uni-siegen.de.

# 1 The problem

International environmental agreements (IEAs) are essential for the stabilization of the world climate at safe levels through the effective reduction of global carbon emissions. The first legally binding international agreement on climate protection, the Kyoto Protocol, has been criticized because it includes commitments for a small number of countries only and is therefore likely to accomplish very little in terms of global emission reduction (Buchner et al. 2002). It will run out in 2012, and the prospects are bleak for reaching a new IEA which accomplishes both attracting many signatories and reducing global emissions significantly. The tedious practical negotiations and the serious global change challenge call for continued investigations of the theoretical foundations of successful and effective IEAs.

Since the early 1990s an economic literature has developed on self-enforcing IEAs. An IEA is said to be self-enforcing or stable if no signatory country has an incentive to leave the IEA and no non-signatory country has an incentive to join the IEA. The seminal papers on self-enforcing IEAs include Barrett (1992, 1994), Hoel (1992) and Carraro and Siniscalco (1993). Most papers are quite pessimistic about the stability of large IEAs. Carraro and Siniscalco (1991), Hoel (1992) and Finus (2001) find that a stable IEA consists of three countries when the climate damage is linear and of two countries when the climate damage is quadratic. These papers assume that both signatories and non-signatories behave in a Cournot-Nash fashion.

Another strand of the IEA literature which we will follow in the present paper makes use of the Stackelberg assumption portraying the climate coalition[1] as Stackelberg leader and all non-cooperative countries as Stackelberg followers. In that framework Barrett's (1994) simulation results suggest the existence of stable coalition sizes between two and the grand coalition. However, Diamantoudi and Sartzetakis (2006) and Rubio and Ulph (2006) proved that large stable IEAs imply zero emissions (corner solutions) or negative emissions. Negative emissions must clearly be ruled out in models without stock pollution because it is infeasible to abate more emissions than are generated. Therefore, the approach of Rubio and Ulph (2006) is correct to introduce non-negativity constraints for (net) emissions which then generates zero-emission corner solutions under certain parameter constellations. As Rubio and Ulph point out the reason for such corner solutions is the assumption of non-essential emissions which is standard in the literature on IEAs. Emissions are non-essential when the marginal benefit from emissions is positive but finite for zero emissions. That assumption may be plausible for some pollutants, e.g. for CFC emissions in the context of the ozone

[1] In the present paper we use the terms IEA and (climate) coalition as synonyms because our exclusive focus is on a single coalition.

problem, but less so for carbon emissions in the context of climate change which is the focus of the present paper. Hence the first-best strategy would be assuming emissions (or rather the consumption of fossil energy) to be essential. However, for reasons of tractability and comparability with pertaining literature we follow Diamantoudi and Sartzetakis (2006) and restrict parameter values to ensure that the resultant emissions are always strictly positive. Under that constraint (along with the assumption of non-essential emissions) Diamantoudi and Sartzetakis (2006) as well as Rubio and Ulph (2006) find that the number of signatories of self-enforcing IEAs is not larger than four. Hence large self-enforcing IEAs cannot be expected under the Stackelberg assumption.

The *basic model* of an IEA employed by Barrett (1994), Diamantoudi and Sartzetakis (2006) and Rubio and Ulph (2006) and others is a static model of symmetric countries where each country's domestic emissions generate domestic welfare that is decreasing at the margin and where all countries' emissions create a welfare loss which is uniform across countries and increasing at the margin.[2] That model does not account for production, consumption, markets and international trade and thus captures the world economy in a rudimentary way only. It has been extended in various directions (Finus 2003).[3] For example, Hoel and Schneider (1997) introduce transfer schemes in the coalition formation process, Kolstad (2007) studies systematic uncertainty and Carbone et al. (2009) use the basic model for an empirical investigation of how international emission trading impacts on IEAs. However, we are not aware of studies on the formation of IEAs[4] that model in more detail the economies of individual countries and their economic interdependencies.

The present paper aims to extend the basic model along these lines and then investigates the impact of that extension on the stability of IEAs in the Stackelberg leader-follower framework. We will add structure to the national economies by introducing a consumer good and fossil fuel that are produced in each country, consumed by its representative consumer and traded on world markets.[5] In this general equilibrium framework we first briefly

[2]Barrett (1994) formalizes abatement and therefore his approach seems to differ from the basic model, at first glance. However, as pointed out by Diamantoudi and Sartzetakis (2006, Section 4), Barrett's model is equivalent to the *basic model* as long as abatement does not exceed the flow of emissions.

[3]Modifying and extending, respectively, the basic model Barrett (1999) and Hannesson (2010) show that stable coalitions may consist of a large number of countries presupposed the coalition countries behave as a Nash player.

[4]There are also studies relaxing the assumption of the basic model that countries are identical (e.g. Barrett 2001). In the present paper we will stick to that assumption to keep our model tractable.

[5]Despite the importance of international trade for the formation of IEAs, to our knowledge there is only one paper dealing with that issue, and that is Barrett (1997) who illustrates in a partial equilibrium model with abatement how trade policy may help support stable IEAs. Copeland and Taylor (2005) study the role of international trade in a model of non-cooperative heterogenous countries coping with a global (climate) externality. They do not address the formation of coalitions, however.

characterize the business-as-usual scenario where the governments of all countries play Nash maximizing domestic welfare by choosing their emission caps as best replies to the other countries' emission caps. We then turn to our central theme, the characterization of self-enforcing IEAs in the Stackelberg model.

For the case that signatories act as Stackelberg leader and equilibrium emissions are positive, we find that - depending on parameter constellations - international trade may significantly increase the size of stable IEAs. That is, the conditions for successful sub-global cooperative action appear to be more favorable than suggested by the *basic model* of the IEA literature referred to above. Unfortunately, the hope for a more optimistic view on *effective* cooperative emission reductions turns out to be unwarranted because our second main finding is that if an IEA of any size is self-enforcing, the corresponding allocation of world emissions is approximately the same as in its absence, i.e. in business as usual (BAU). We hasten to add that these results are obtained in a very simple model making use of parametric functions and numerical examples. It is inappropriate, therefore, to take them as reliable indicators for the outcome of the highly complex ongoing international climate negotiations. Nonetheless, they provide some support for the disturbing view that any attempt to form a sub-global climate coalition (of whatever size) is futile.

As the introduction of international trade represents a major extension of the IEA literature it is natural to highlight its impact on results by looking at the outcome of our model in the absence of international trade (autarky). Somewhat unexpectedly, when all countries are autarkic, our model turns out to coincide with the *basic model* of the literature on IEAs which has established, as reported above, that the number of signatories in self-enforcing IEAs is very small. We find that allowing for trade leads to larger stable coalitions than under autarky. Concerning the effectiveness of stable coalitions in the autarky scenario, the outcome is as in the free-trade scenario: The allocation in an equilibrium with a stable coalition is almost the same as in BAU.

The paper is organized as follows. Section 2 introduces the model and briefly analyzes the business-as-usual scenario which serves as a benchmark throughout the paper. Section 3.1 prepares for the analysis of self-enforcing IEAs in Section 3.2 by characterizing the outcome of the Stackelberg game and, in particular, its dependence on the exogenously given size of coalitions. Section 4 deals with the role of international trade for the results by comparing the regimes of free trade and autarky and by linking the case of autarky to the basic model of the coalition formation literature. Section 5 concludes.

## 2 The model

The world economy consists of $n$ identical countries. Each country produces two consumer goods. The first is a standard composite good, called *good X* (quantity $x_i$) and the second is a fossil energy carrier (quantity $e_i$), e.g. oil, gas or coal extracted from domestic fossil reserves. We refer to that good simply as *fuel*.[6] Each country's production technology is represented by the production possibility frontier[7]

$$x_i^s = T(e_i^s) \quad i = 1, \ldots, n, \tag{1}$$

where the function $T$ is decreasing and strictly concave in $e_i^s$. The transformation function (1) implies that both commodities are produced by means of domestic productive factors (e.g. labor and capital) whose endowments are given. The utility[8]

$$V(e_i^d) + x_i^d - D\left(\sum_j e_j^d\right) \tag{2}$$

of the representative consumer of country $i$ is additive separable in all arguments and linear in the consumption $x_i^d$ of good $X$. $V$ is increasing and concave, and $D$ is increasing and convex in its argument. The consumption of fuel generates the greenhouse gas carbon dioxide whose emission is proportional to fuel consumption. Emission units are chosen such that $e_i^d$ denotes both fuel demanded by consumer $i$ and carbon emissions from burning fuel. There is no abatement technology for emission reduction.[9] The function $D$ captures the climate damage caused by worldwide carbon emissions from burning fuel.

For the sake of more specific results, throughout the paper we will specify the functions $T$, $V$ and $D$ from (1) and (2) by the following quadratic functional forms:[10]

$$T(e_i^s) = \bar{x} - \frac{\alpha}{2}\left(e_i^s\right)^2, \quad V(e_i^d) = ae_i^d - \frac{b}{2}(e_i^d)^2, \quad D\left(\sum_j e_j^d\right) = \frac{1}{2}\left(\sum_j e_j^d\right)^2, \tag{3}$$

[6]Households do not consume fuel directly but use fuel as input in a linear household production function to produce e.g. the commodities heat or transportation services. Throughout the rest of the paper we leave off the household production technology and interpret fuel as consumer good.

[7]The superscript $s$ indicates quantities supplied. Upper-case letters are reserved to denote functions. Subscripts attached to them indicate partial derivatives.

[8]The superscript $d$ indicates quantities demanded.

[9]Carbon capture and sequestration is a potential abatement technology which is unlikely to be applied on a large scale in the near or medium term future.

[10]In (3) the parametric form of $T(e_i^s)$ can be 'microfounded' as follows. Let $\bar{r}$ be country $i$'s endowment of a (composite) production factor and consider the production functions $x = \alpha_x r_x$ and $e = (r_e/\alpha_e)^{1/2}$ with $r_e + r_x = \bar{r}$. $\alpha_e, \alpha_x$ are positive constants. The quadratic transformation function in (3) is straightforward from these three equations when setting $\bar{x} := \alpha_x \bar{r}$ and $\alpha := \alpha_x \alpha_e$.

where $\bar{x}$, $a$, $b$ and $\alpha$ are positive parameters.

Our stylized model (1) and (2) of the individual country's economy neglects fuel as an intermediary input in the production of good $X$. All fuel goes from production directly to consumers where 'fuel production' can be interpreted to include extraction of fossil energy carriers as well as production of electricity, gasoline, gas or coal for non-business usage.[11] Although in practice climate regulation does not only apply to the consumers' energy demand but also to energy-consuming industries, as e.g. in the EU emission trading scheme, we maintain that our simplification still captures the central issue of emission regulation. Whether fuel consumption of industries or of consumers is regulated, in both cases more stringent emission caps require raising the domestic price for fuel consumption which, in turn, induces allocative displacement effects via changes in relative prices.

There are perfectly competitive world markets for good $X$ (price $p_x \equiv 1$) and for fuel (producer price $p$), and the markets are in equilibrium if

$$\sum_j x_j^s = \sum_j x_j^d \quad \text{and} \quad \sum_j e_j^s = \sum_j e_j^d. \tag{4}$$

The firms' supply of fuel is straightforward. Taking prices as given, the (aggregate) producer $i$ maximizes profits $x_i^s + pe_i^s$ subject to (1) which yields the first-order condition

$$p = -T'(e_i^s) \quad \text{for} \quad i = 1, \ldots, n. \tag{5}$$

Combined with (1), equation (5) implies a fuel supply function

$$e_i^s = E^s(p) \quad \text{with} \quad E_p^s > 0 \quad \text{for} \quad i = 1, \ldots, n. \tag{6}$$

Each government $i$ regulates domestic carbon emissions by enforcing an emission cap $e_i$. For the time being we suppose these caps are arbitrarily fixed and tight enough to be binding. To implement its emission cap, government $i$ issues the amount $e_i$ of emission permits and auctions them at the permit price $\pi_i$. Consumers in country $i$ need to acquire emission permits to match their purchase of fuel. The representative consumer $i$ ignores the impact of her emissions on climate damage and maximizes her (consumption) utility $V(e_i^d) + x_i^d$ subject to her budget constraint

$$x_i^d + (p + \pi_i)e_i^d = y_i, \quad \text{where} \quad y_i := x_i^s + pe_i^s + \pi_i e_i^d \tag{7}$$

is consumer $i$'s income (= profit income plus recycled revenues from the permit auction). From the first-order condition $p + \pi_i = V'(e_i^d)$ follows a fuel demand function

$$e_i^d = E^d(p + \pi_i) \quad \text{for} \quad i = 1, \ldots, n. \tag{8}$$

[11] Such simplifications are driven by limits of tractability. We also wish to recall, however, that the model of the present paper is far more complex than the basic model of IEA (e.g. Finus 2003, Section 2.3) which does without specifying production, consumptions and markets, as we have pointed out in the introduction.

The result of auctioning the permits obviously is

$$e_i^d = e_i \quad \text{for} \quad i = 1, \ldots, n. \tag{9}$$

Combining the equilibrium condition $\sum_j e_j^s = \sum_j e_j^d$ from (4) with (6) and (9) yields

$$e_i^s = E^s(p) = \frac{\sum_j e_j}{n} \quad \text{for} \quad i = 1, \ldots, n. \tag{10}$$

Equation (10) determines the unique equilibrium price of fuel and also establishes that in equilibrium all firms produce the same amount of fuel, $\sum_j e_j/n$. From (5), (8) and (9) follows $e_i = E^d\left[-T'\left(\frac{\sum_j e_j}{n}\right) + \pi_i\right]$. This equation determines the unique equilibrium permit price. The equilibrium supplies and demands on the market for good $X$ are

$$x_i^s = T\left(\frac{\sum_j e_j}{n}\right) \quad \text{and} \quad x_i^d = T\left(\frac{\sum_j e_j}{n}\right) - T'\left(\frac{\sum_j e_j}{n}\right)\left(\frac{\sum_j e_j}{n} - e_i\right), \tag{11}$$

where the first equation in (11) is implied by (1) and (10) and the second by (1), (7), (9) and (10). It readily follows from (11) that the market for good $X$ is in equilibrium, if the fuel market is in equilibrium.

To sum up, in the world economy with non-cooperative emission cap regulation there is a unique competitive equilibrium for every profile $(e_1, \ldots, e_n)$ of binding emission caps. That is, in equilibrium all demands and supplies, $e_i^s$, $e_i^d$, $x_i^s$, $x_i^d$, $i = 1, \ldots, n$, and the prices $p$ and $\pi_i$, $i = 1, \ldots, n$, are determined by $(e_1, \ldots, e_n)$. Combining welfare (2) with (9), (10) and (11) results in the equilibrium welfare of country $i = 1, \ldots, n$,

$$W^i(e_1, \ldots, e_n) := V(e_i) + T\left(\frac{\sum_j e_j}{n}\right) - \left(\frac{\sum_j e_j}{n} - e_i\right) T'\left(\frac{\sum_j e_j}{n}\right) - D\left(\sum_j e_j\right). \tag{12}$$

So far we have considered governments that fix national emission caps in an arbitrary way. From now on their objective function is supposed to be national welfare, (12). Before addressing the case of cooperation in emission regulation we briefly investigate the benchmark case of global non-cooperation. In game-theoretic language, the $n$ governments are the players of a non-cooperative game. Their strategies are national emission caps and their payoff functions are national welfares $W^i(e_1, \ldots, e_n)$ from (12). The natural solution concept is the Nash equilibrium, a state where each government's emission cap is the best response to each other government's emission cap. As usual, we refer to that equilibrium as business as usual (BAU). In terms of the formal model, government $i$ chooses that cap $e_i$ which maximizes $W^i(e_1, \ldots, e_n)$ for given caps $(e_1, \ldots, e_{i-1}, e_{i+1}, \ldots, e_n)$. Differentiation of (12) with respect to $e_i$ yields the first-order condition[12]

$$W_{e_i}^i = V'(e_i) + T'\left(\frac{\sum_j e_j}{n}\right) - \frac{1}{n}\left(\frac{\sum_j e_j}{n} - e_i\right) T''\left(\frac{\sum_j e_j}{n}\right) - D'\left(\sum_j e_j\right) = 0 \tag{13}$$

[12]Throughout the rest of the paper we restrict our attention to interior solutions.

for $i = 1, \ldots, n$. Eichner and Pethig (2012) show that (13) is equivalent to a best reply function $\tilde{R}$ satisfying

$$e_i = \tilde{R}\left(\sum_{j \neq i} e_j\right) \tag{14}$$

whose first derivative is in the interval $]-1, 0[$ under mild restrictions. Hence there exists a unique symmetric Nash equilibrium satisfying $e_i = e_j$ for all $j \neq i$. Another immediate consequence of symmetry is that international trade does not take place in equilibrium. When these observations are considered in (13), the uniform Nash equilibrium cap, denoted $e_o$, is implicitly determined by

$$V'(e_o) + T'(e_o) - D'(ne_o) = 0. \tag{15}$$

Inserting the parametric functions (3) in (15) readily yields $e_o = \frac{a}{\alpha+b+n}$. The allocation rule (15) shows that each country sets its BAU emission cap $e_o$ such that its marginal benefit of consumption, $V'(e_o) + T'(e_o)$, equals its marginal climate damage, $D'(ne_o)$. If the countries would disregard their own impact on climate damage (i.e. if we would drop the term $D'(ne_o)$ in (15)) national equilibrium emissions would exceed $e_o$. Hence in BAU some emission reduction is in the countries' self-interest. It is also clear that total emissions $ne_o$ in BAU exceed total emissions in the optimal fully cooperative solution, since all countries disregard in BAU the positive external effects of their emission reduction on the other countries.

# 3 Climate coalition as Stackelberg leader

Suppose now that some countries are members in a given cooperative climate coalition, whereas all other countries continue to act non-cooperatively. For the purpose of the formal analysis we lump together the first $m$ countries, $2 \leq m < n$, in one group, denoted group $C := \{1, 2, \ldots, m\}$ with $C$ for coalition, and collect all remaining countries in another group, denoted group $F := \{m+1, \ldots, n\}$ with $F$ for fringe. Our focus will be on a game of sequential choice of emission caps in which the coalition is the Stackelberg leader and moves first and the fringe countries are Stackelberg followers. The coalition formation literature has made ample use of the Stackelberg assumption (Finus 2001) and we refer the reader to that literature for information on the discussion about the plausibility and relative merits of the Nash concept on the one hand and the Stackelberg concept on the other.[13] Our aim is to investigate how the Stackelberg assumption drives the outcome of the game when we extend the basic model as outlined in Section 2.

[13] Eichner and Pethig (2012) is a companion paper where we model the climate coalition as a Nash player.

## 3.1 Climate coalitions and coalition sizes

In the present section we aim to characterize the allocation of the Stackelberg equilibrium (to be specified below) for alternatively given coalition sizes and thus prepare for the analysis of coalition stability in the next Section 3.2. The objective of the climate coalition $C$ is to maximize the joint welfare $\sum_{j\in C} W^j(e_1,\ldots,e_n)$ of its members taking the behavior of the fringe countries into account. Since all coalition countries are alike, $e_i = e_j$ for all $i,j \in C$ is a necessary maximum condition which allows us to set $e_i = e_c$ for all $i \in C$. Thus the coalition can be treated as a single player whose strategy will be denoted as $s_c := me_c$. We continue portraying fringe countries as non-cooperative Nash players, and therefore (13) still applies for each fringe country. As (13) cannot be satisfied for $i,j \in F, i \neq j$, unless $e_i = e_j$, we proceed by setting $e_i = e_f$ for all $i \in F$. With this notation, each fringe country's best-reply function (14) reads $e_f = \tilde{R}[s_c + (n-m-1)e_f]$, and Eichner and Pethig (2012) show that this equation implies a function $R$ satisfying $(n-m)e_f = R(me_c, m)$ or

$$s_f = R(s_c, m) \quad \text{with} \quad R_{s_c} \in ]-1, 0[, \tag{16}$$

where $s_c := me_c$ and $s_f := (n-m)e_f$.

According to (16) the fringe countries can be treated as if they act as a single player whose strategy is $s_f$. In that sense $R$ is the 'aggregate' best reply function of 'the fringe'. However, it is important to emphasize that, by construction, (16) does not imply any cooperation among fringe countries. Although the function $R$ is a purely formal transformation of $\tilde{R}$ from (14), it turns out to be an important analytical tool. In the subsequent analysis we make use of $R$ and its specific properties summarized in[14]

**Lemma 1.** *The function $R$ satisfies $\hat{s}_c := R^{-1}[(s_f = 0, m] > 0$ for all $m \in ]0, n[$, $R_m(s_c, m) < 0$ for all $s_c < \hat{s}_c$, all $m \in ]0, n[$, $R_{s_c s_c} = 0$ and $R_{s_c m} > 0$.*

Thus, the graph of the best-reply function $R$ is a negatively sloped straight line. Its point of intersection with the $s_c$ axis, $\hat{s_c}$, is independent of $m$ and it rotates around that point towards [away from] the origin, if $m$ increases [decreases]. Making use of the newly introduced notation $s_f := (n-m)e_f$, we next express total emissions as $\sum e_j = s_c + s_f$ and rewrite the welfare of individual countries, (12), as

$$W^c(s_c, s_f, m) := V\left(\frac{s_c}{m}\right) + T\left(\frac{s_c + s_f}{n}\right) - \left(\frac{s_c + s_f}{n} - \frac{s_c}{m}\right) T'\left(\frac{s_c + s_f}{n}\right) - D(s_c + s_f) \tag{17}$$

[14] The proof of Lemma 1 is delegated to the Appendix A.

for all countries in group $C$ and as

$$W^f(s_c, s_f, m) := V\left(\frac{s_f}{n-m}\right) + T\left(\frac{s_c+s_f}{n}\right) - \left(\frac{s_c+s_f}{n} - \frac{s_f}{n-m}\right) T'\left(\frac{s_c+s_f}{n}\right) - D(s_c+s_f) \tag{18}$$

for all countries in group $F$. For convenience of notation and later reference we refer to $(-D(s_c+s_f))$ as the *climate welfare* of an individual country and to $W^j(s_c, s_f, m)+D(s_c+s_f)$ as the *consumption welfare* of countries in the coalition $(j = c)$ or the fringe $(j = f)$.

Being the Stackelberg leader the coalition with a given number $m \in \{1, \ldots, n\}$ of member countries accounts for (16) such that its objective function is the aggregate welfare $mW^c[s_c, R(s_c, m), m]$. The fringe countries observe the leader's action $s_c$. Their 'aggregate' response is $s_f = R(s_c, m)$ and the resultant welfare is $W^f(s_c, R(s_c, m), m)$ for each individual fringe country. How $W^c(\cdot)$ and $W^f(\cdot)$ depend on $s_c$ is specified in[15]

**Lemma 2.** *$W^c(\cdot)$ is inverse u-shaped and strictly concave in $s_c$, $\left(\frac{\mathrm{d}^2 W^c}{\mathrm{d}s_c^2} < 0\right)$, and $W^f(\cdot)$ is strictly decreasing in $s_c$, $\left(\frac{\mathrm{d}W^f}{\mathrm{d}s_c} < 0\right)$.*

$\mathrm{d}W^f/\mathrm{d}s_c < 0$ conforms to intuition because the larger is the coalition's contribution to climate damage the greater is the fringe countries' aggregate effort to reduce their emissions. To understand the dependence of $W^c(\cdot)$ from $s_c$ suppose that $s_c = R(s_c, m)$ is very small initially. Then the climate welfare of coalition countries is high but their consumption welfare is low due to their high mitigation efforts. Increasing $s_c$ subject to (16) raises the consumption welfare by more than it reduces the climate value. The opposite effects are created, if $s_c = R(s_c, m)$ is very large initially and is then successively reduced. Technically, speaking, the property $\mathrm{d}^2 W^c/\mathrm{d}s_c^2 < 0$ secures a unique solution to the coalition's optimization problem

$$\max_{s_c \in [0, mT^{-1}(0)]} mW^c[s_c, R(s_c, m), m]. \tag{19}$$

The Stackelberg equilibrium is the solution of (19), denoted $s_c^*$. It is implicitly defined by the marginal condition

$$\frac{W^c_{s_c}(s_c^*, s_f^*, m)}{W^c_{s_f}(s_c^*, s_f^*, m)} = -R_{s_c}(s_c^*, m), \tag{20}$$

where $s_f^* = R(s_c^*, m)$. The equilibrium condition (20) is the standard representation of a Stackelberg equilibrium as a point in the strategy space, here $(s_c^*, s_f^*)$, where the best-reply function $R$ of the fringe and an iso-welfare curve of the coalition are tangent.

In the sequel we will characterize the solution of (19), its relation to the BAU equilibrium and its dependence on the (exogenous) size of the coalition. We proceed in several steps beginning with the implications of an arbitrary action $s_c \in [0, mT^{-1}(0)]$ of the leader.

[15] Lemma 2 is proven in the Appendix B.

**The coalition's anticipation of the fringe's reactions as driving force of outcomes**
The best-reply function of the fringe, (16), is of special interest because all feasible outcomes necessarily satisfy that function. Accounting for $R$ the coalition knows that its own emissions and those of the fringe are strategic substitutes and so it takes into consideration that if it chooses the cap $s_c$ total emissions will be

$$s_c + s_f = s_c(1 + R_{s_c}) + R(0). \tag{21}$$

Note that $(1 + R_{s_c}) \in ]0, 1[$ because $R_{s_c} \in ]-1, 0[$. From equation (21) readily follows $\partial(s_c + s_f)/\partial s_c = (1 + R_{s_c}) \in ]0, 1[$ which means that if the coalition reduces its emissions, $\Delta s_c < 0$, [increases its emissions, $\Delta s_c > 0$], total emissions will shrink [expand], but by less than $\left|\Delta s_c\right|$. In the climate change literature this phenomenon is referred to as carbon leakage for the case $\Delta s_c < 0$. The leakage rate is usually expressed by $\left|R_{s_c}\right|$. Since $\left|R_{s_c}\right| \in ]0, 1[$, a leakage rate greater than one, the so-called 'green paradox', does not occur in our model. Since $R_{s_c m} > 0$ has been established in Lemma 1, the leakage rate is declining in the coalition size - which conforms to intuition. As an implication of (21) we get[16]

$$e_c \gtreqless e_o \quad \Longleftrightarrow \quad s_c + s_f \gtreqless n e_o, \tag{22}$$

because the frince countries' response $\Delta e_f > 0[\Delta e_f < 0]$ to the coalition countries' action $\Delta e_c < 0[\Delta e_c > 0]$ does not fully compensate the action $\Delta e_c$. Moreover, from $R_{s_c} < 0$ and $x_i^s = T\left(\frac{s_c + s_f}{n}\right)$ for $i = 1, \ldots, n$, follows

$$e_c \gtreqless e_o \quad \Longleftrightarrow \quad \text{coalition} \left\{ \begin{array}{c} \text{imports} \\ \text{doesn't trade} \\ \text{exports} \end{array} \right\} \text{fuel.} \tag{23}$$

If $e_c - e_o > 0$, the fringe countries' response is $e_f < e_c$. Hence $e_f < (s_c + s_f)/n < e_c$. As the supply of fuel, $(s_c + s_f)/n$, is the same across all countries, the coalition imports fuel. An analogous argument applies for $e_c - e_o < 0$ which explains (23). The total differential[17] of (21) reads

$$\mathrm{d}(s_c + s_f) = \left[ \underbrace{(1 + R_{s_c})e_c + s_c R_{s_c m}}_{(+)} \right] \cdot \mathrm{d}m + \underbrace{m(1 + R_{s_c})}_{(+)} \cdot \mathrm{d}e_c. \tag{24}$$

(24) shows that if $e_c$ is kept constant, total emission are rising in $m$ because the fringe's responding emission reduction falls short of the coalition's emission increase, $e_c \mathrm{d}m$. While

[16] More precisely, (21) implies $s_c + s_f - n e_o = s_c(1 + R_{s_c}) + R(0) - m e_o(1 + R_{s_c}) - R(0) = m(1 + R_{s_c})(e_c - e_o)$ and hence (22).

[17] It is analytically convenient to treat $m$ as a real number in $[1, n]$ although we will keep in mind that $m \in \{1, \ldots, n\}$ for real-world coalitions.

$\partial(s_c + s_f)/\partial e_c > 0$ is a generalized statement of (22), the obvious observation is that the magnitude of $\partial(s_c + s_f)/\partial e_c$ depends on the coalition size $m$:

$$\frac{\partial^2(s_c + s_f)}{\partial e_c \partial m} = (1 + R_{s_c}) + mR_{s_c m} = \frac{\{[1 - (n - m - 1)R_{s_c}]^2 - mR_{s_c}\}(1 + R_{s_c})}{[1 - (n - m - 1)R_{s_c}]^2} > 0. \quad (25)$$

According to (25) the impact of variations in $e_c$ on total emissions $s_c + s_f$ is increasing in $m$, $\partial(s_c + s_f)/(\partial e_c \partial m) > 0$. That is, if the coalition tightens [relaxes] the cap $e_c$ for all members by one unit, the reduction [increase] in total emissions is the larger, the larger is the coalition. Large coalitions are more effective in curbing total emissions because the leakage rate is declining in the coalition size.

**Allocation resulting from the coalition's (not necessarily optimal) choice** $e_c \neq e_o$ The case $e_c > e_o$ is illustrated in Figure 1 that contains the production possibility curve $AB$ and several consumption welfare indifference curves denoted by $u$. All these curves are the same for all countries. The point $P_o$ on the transformation curve is assumed to be the BAU equilibrium point characterizing all countries' production and consumption in BAU ($x^s_{co} = x^s_{fo} = x^d_{co} = x^d_{fo} = x_o$ and $e^s_{co} = e^s_{fo} = e_{co} = e_{fo} = e_o$).[18] As $e_c > e_o$ implies $\frac{s_c+s_f}{n} > e_o$ according to (22) and $e^s_c = e^s_f = \frac{s_c+s_f}{n}$, we find that $T\left(\frac{s_c+s_f}{n}\right) < T(e_o) = x_o$. Hence the production point lies on the curve $AB$ to the right of point $P_o$, marked as point $P$ in Figure 1. The straight line $EF$ in Figure 1 is tangent to the transformation curve in point $P$ and therefore represents the terms of trade, $\tan\alpha = p$. The consumption points of all coalition and fringe countries must lie on $EF$. According to (22) $e_c > e_o$ implies that the coalition countries import fuel and export good $X$. Therefore the coalition countries' [fringe countries'] consumption point $K_c[K_f]$ is located to the right [left] of the production point $P$. The coalition countries' total welfare is

$$W^c(s_c, s_f, m) = \underbrace{V(e_o) + T(e_o) - D(e_o) +}_{\text{Equilibrium welfare in BAU}}$$

$$+ \overbrace{\underbrace{V(e_c) - V(e_o)}_{(+)} + \underbrace{T\left(\frac{s_c + s_f}{n}\right) - T(e_o)}_{(-)} + \underbrace{\left(e_c - \frac{s_c + s_f}{n}\right) T'\left(\frac{s_c + s_f}{n}\right)}_{(-)}}^{\text{Gain in consumption welfare}} + \overbrace{\underbrace{[D(e_o) - D(s_c + s_f)]}_{(-)}}^{\text{Loss in climate welfare}}.$$

This decomposition of welfare effects demonstrates that Figure 1 is inconclusive regarding the coalition countries' change in total welfare compared to BAU. As drawn in Figure 1, the fringe countries suffer a loss in consumption welfare ($u_f < u_o$), and a loss in climate welfare, the same as suffered by the coalition countries, such that the fringe countries lose on both accounts compared with BAU.

[18] Recall from Section 2 that the superscript $s$ stands for commodities supplied not to be confounded with the strategies $s_c := me_c$ and $s_f := (n - m)e_f$.

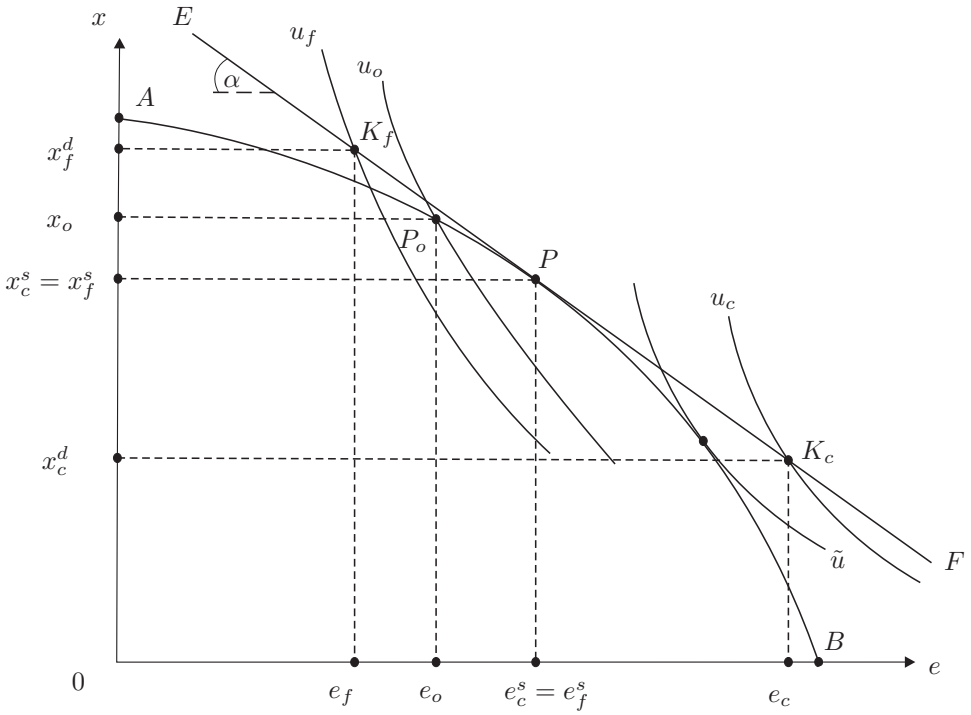


Figure 1: Allocative impact of the coalition's strategy $s_c = me_c > me_o$

Next consider the coalition's action $e_c < e_o$ illustrated in Figure 2. Since the arguments are analogous to those in the preceding paragraph, it suffices to summarize the results. In case of $e_c < e_o$ the production point $P$ corresponding to $[me_c, R(me_c)]$ lies on the transformation curve to the right of the BAU point $P_o$ and the coalition countries export fuel. Hence these countries suffer a consumption welfare loss ($u_c < u_o$) while in the fringe countries enjoy an increase in their consumption welfare ($u_f > u_o$). For the coalition countries the total welfare change of moving from $e_o$ to $e_c < e_o$ is the result of two opposite partial welfare effects, as in the case $e_c > e_o$ dealt with above.

**Coalition size, equilibrium emissions and welfares, and their relation to BAU** The preceding discussion of the Figures 1 and 2 served to illustrate the consequences of the choice of some $s_c \neq me_o$ on the part of the coalition. However, these figures provide no information about whether the scenarios they illustrate can be the result of the coalition's optimal choice $s_c^*$ and if so, under which conditions. We now address that issue focussing on the coalition size as a determinant of the coalition's optimal choice strategy. It is clear from our previous analysis that $e_c^*$ and with it the entire Stackelberg equilibrium allocation are uniquely determined by - and vary with - the coalition size. To formalize that observation

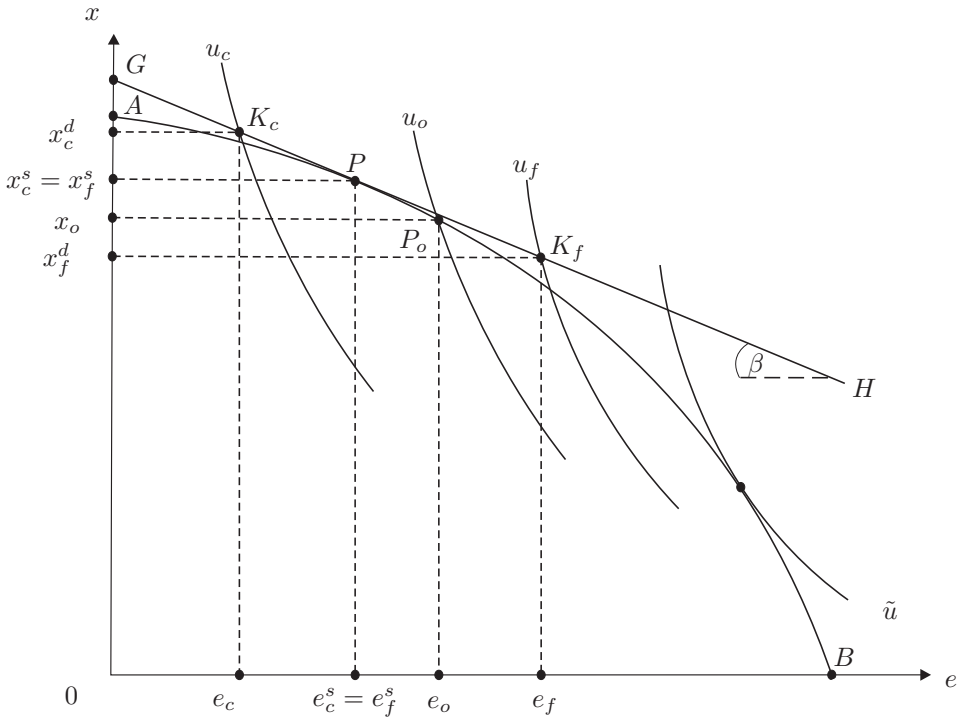


Figure 2: Allocative impact of the coalition's strategy $s_c = me_c < me_o$

it is convenient to introduce the notation

$$e_c^* = \mathcal{E}^c(m), \quad e_f^* = \mathcal{E}^f(m),$$
$$\mathcal{W}^c(m) := W^c[m\mathcal{E}^c(m), (n-m)\mathcal{E}^f(m), m] \quad \text{and}$$
$$\mathcal{W}^f(m) := W^f[m\mathcal{E}^c(m), (n-m)\mathcal{E}^f(m), m].$$

A first but important step toward answering the questions raised above takes[19]

**Proposition 1.** *For analytical convenience consider the interval* $[1, n]$ *to be the domain of coalition sizes. The Stackelberg equilibrium associated with the coalition of size* $\tilde{m} \in [1, n]$ *coincides with the non-cooperative BAU equilibrium, if and only if*

$$\tilde{m} := \frac{(\alpha + b + n)n^2}{\alpha(2n-1) + n^2(b+1)} > 1. \tag{26}$$

Proposition 1 specifies the link between Stackelberg equilibria and the non-cooperative BAU equilibrium. For the coalition it is optimal to choose the BAU emissions $e_c^* = e_o$ (leading to $e_f^* = e_o$), if and only if it has $\tilde{m}$ members. $\mathcal{E}^c(m) \neq e_o$ and $\mathcal{W}^c(m) \geq \mathcal{W}^c(\tilde{m})$ for all $m \neq \tilde{m}$ follows immediately from the observations that the benchmark coalition size $\tilde{m}$ is unique and that for any given $m$ the coalition can always choose the emission cap $e_c = e_o$ which then leads to the BAU equilibrium. According to (26) $\tilde{m}$ varies with the model parameters and that feature will turn out to be of special interest below.

[19]The proof of Proposition 1 can be found in the Appendix C.

Taking into account that in the real world the number of coalition members $m$ must be an integer in $\{1,\ldots,n\}$ the real message of Proposition 1 is that BAU and Stackelberg equilibria do not coincide unless $\tilde{m}$ happens to be an integer. If it is not, define $\tilde{m}^{(-)}[\tilde{m}^{(+)}]$ as the largest [smallest] integer smaller than [larger than] $\tilde{m}$ and observe that the Stackelberg equilibria attained by coalitions of size $\tilde{m}^{(-)}$ and $\tilde{m}^{(+)}$, respectively, come closer to the BAU equilibrium than any Stackelberg equilibrium of coalitions with $m < \tilde{m}^{(-)}$ or $m > \tilde{m}^{(+)}$ members. In that sense we say that in coalitions of size $\tilde{m}^{(-)}$ or $\tilde{m}^{(+)}$ the Stackelberg equilibrium is approximately equal to the BAU outcome.

With the coalition size $\tilde{m}$ as a benchmark we are able to shed more light on the links between coalition size and deviations from BAU of emissions and welfare levels in Stackelberg equilibria. Suppose, the coalition of size $m \in [1,n[$ chooses the strategy $s_c = me_o$ and thus implements the BAU equilibrium. For all coalitions of size $m \neq \tilde{m}$ the strategy $s_c = me_o$ is clearly feasible but sub-optimal. Hence $MWC_o(m) \neq 0$ for all $m \neq \tilde{m}$, where $MWC_o(m)$ is a shorthand for the "**M**arginal (aggregate) **W**elfare of a **C**oalition of size $m \neq \tilde{m}$ evaluated at the 'BAU equilibrium strategy' $s_c = me_o$ ". We prove in the Appendix C that[20]

$$MWC_o(m) \gtreqless 0 \quad \Longleftrightarrow \quad m \lesseqgtr \tilde{m}. \tag{27}$$

For the interpretation of (27) we invoke our result from the proof of (27) in the Appendix that the coalition's marginal *consumption* welfare in BAU is independent of the coalition size, so that variations in *total* marginal welfare result exclusively from variations in the coalition's marginal *climate* welfare. Hence, total BAU emissions $ne_o$ are considered as too large by large coalitions ($m > \tilde{m}$) and as too small by small coalitions ($m < \tilde{m}$).[21] Thus combining the information of (27) with the properties of $W^c[me_c, R(me_c,m), m]$ in Lemma 2 we get

$$\mathcal{E}^c(m) \gtreqless e_o \quad \Longleftrightarrow \quad m \lesseqgtr \tilde{m}. \tag{28}$$

In view of (28) the scenario depicted in Figure 1 [Figure 2] can be taken as an illustration of the Stackelberg equilibrium with coalition size $m < \tilde{m}$ [$m > \tilde{m}$]. The rationale of (28) is straightforward from (25). In case of $m < \tilde{m}$ the leakage rate is high so that the coalition would achieve only a small reduction in total emissions, if it reduces $e_c$ by some increment, say $\Delta e_c > 0$, to $e_c - \Delta e_c$. Reducing total emissions would be very expensive. If, instead, the coalition adds rather than subtracts $\Delta e_c$ to the cap $e_c$, the resulting increase in total emissions is small because of the high leakage rate, but the gain in consumption welfare is relatively large. Mirror symmetric arguments apply to the case $m > \tilde{m}$. Since according to

[20] Throughout the paper the subscript "$o$" refers to the BAU equilibrium.

[21] The reason for that differential effect is (25) according to which the effectiveness of curbing total emissions is increasing in the coalition size because the leakage rate declines with increasing coalition size.

(21) and (22) leakage rates are always less than unity,

$$[m\mathcal{E}^c(m) + (n-m)\mathcal{E}^f(m)] \gtreqless ne_o \quad \Longleftrightarrow \quad m \lesseqgtr \tilde{m} \tag{29}$$

follows from (28). These equivalences can also be traced in Figures 1 and 2. For the case $m < \tilde{m}$ (29) says that small coalitions do not mitigate but rather aggravate climate damage: $D[m\mathcal{E}^c(m) + (n-m)\mathcal{E}^f(m)] > D(ne_o)$, if $m < \tilde{m}$. In the scenario of global non-cooperation (BAU) the world suffers from less climate damage than after a small coalition has formed. It would be more appropriate to call such a coalition an 'anti-climate coalition' rather than a 'climate coalition'.

Turning to the coalition country's welfare (27) implies that $\mathcal{W}^c(m)$ is strictly greater than $\mathcal{W}^c(\tilde{m})$ for all $m \neq \tilde{m}$. Hence the function $W^c$ attains its absolute minimum at $m = \tilde{m}$ (but other 'local' minima cannot be excluded here). Moreover, we verify

$$\left.\begin{array}{l} \mathcal{W}^c(m) > W_o > \mathcal{W}^f(m) \\ \mathcal{W}^c(m) = W_o = \mathcal{W}^f(m) \\ \mathcal{W}^f(m) > \mathcal{W}^c(m) > W_o \end{array}\right\} \quad \Longleftrightarrow \quad m \left\{\begin{array}{l} < \\ = \\ > \end{array}\right\} \tilde{m} \tag{30}$$

with $W_o := \mathcal{W}^c(\tilde{m}) = \mathcal{W}^f(\tilde{m})$ as follows. If $m < \tilde{m}$, (28) implies $m\mathcal{E}^c(m) > me_o$ and therefore $\mathcal{W}^f(m) < W_o$ because $\frac{\mathrm{d}W^f}{\mathrm{d}s_c} < 0$ due to Lemma 2. If $m > \tilde{m}$, (28) implies $m\mathcal{E}^c(m) < me_o$ and therefore $\mathcal{W}^f(m) > W_o$ because $\frac{\mathrm{d}W^f}{\mathrm{d}s_c} < 0$ due to Lemma 2. Analogously, $\mathcal{W}^f(m) > \mathcal{W}^c(m)$ for $m > \tilde{m}$ follows from $m\mathcal{E}(m) < me_o$ and Lemma 2 and is also conclusive from Figure 2.

In case of $m < \tilde{m}$ the coalition finds it beneficial to expand own emissions above BAU level which induces the fringe countries to implement more stringent emission caps. The opportunity costs of that climate damage mitigation policy on the part of the fringe countries is consumption foregone. That loss of consumption welfare together with the reduction in climate welfare pushes the fringe countries' total welfare below BAU level. In a sense, the coalition free rides on the fringe countries' mitigation efforts. In case of $m > \tilde{m}$ the roles of both groups are reversed. Now the fringe countries free ride on the coalition's mitigation policy - which is the case that is usually in the focus. As Figure 2 shows, the fringe countries benefit on two margins: Their consumption welfare rises compared to BAU as well as their climate welfare. A general principle appears to be that countries with laxer emission regulation have higher welfare levels. So far, we summarize our results in

**Proposition 2.** *Consider the transition from BAU to the Stackelberg equilibrium. The shift of*

*(i) the coalition country's emissions is characterized in (28);*

*(ii) total emissions is characterized in (29);*

*(iii) the coalition country's and fringe country's welfare is characterized in (30).*

The results of Proposition 2 provide interesting information about the relations between the coalition size, the BAU equilibrium and the Stackelberg equilibrium. However, the functions $\mathcal{E}^h$ and $\mathcal{W}^h$ for $h = c, f$ depend on $m$ in a very complex way such that their curvature cannot be specified analytically. To make progress we resort to a numerical example with the parameter values $a = 1000$, $b = 20$, $\bar{x} = 12$, $\alpha = 1000$ and $n = 10$ which we refer to as Example 1.[22] The Figures 3 and 4 display the pertaining graphs of the functions $\mathcal{E}^h$ and $\mathcal{W}^h$ for $h = c, f$ and the curves of aggregate emissions and welfares, respectively. First observe that (28), (29) and (30) are satisfied in these figures. It is also worth noting that for Example 1 the benchmark coalition size is $\tilde{m} = 4.88$ such that almost 50 % of all countries are members in the coalition whose Stackelberg equilibrium coincides with the BAU equilibrium.[23] The new information conveyed by Example 1 is that the function $\mathcal{E}^c[\mathcal{E}^f]$ is strictly decreasing [increasing] and that $m\mathcal{E}^c + (n-m)\mathcal{E}^f$ is strictly decreasing in $m$.[24] The latter observation is in line with (29) and supplements those equivalences through

$$\left\{\begin{array}{l}\mathcal{E}^c(m) > e_o > \mathcal{E}^f(m)\\ \mathcal{E}^c(m) = e_o = \mathcal{E}^f(m)\\ \mathcal{E}^f(m) > e_o > \mathcal{E}^c(m)\end{array}\right\} \quad \Longleftrightarrow \quad m \left\{\begin{array}{l}<\\ =\\ >\end{array}\right\} \tilde{m}. \tag{31}$$

According to the left panel of Figure 4, the (total) welfare of coalition countries is u-shaped with its unique minimum at $m = \tilde{m}$, whereas $\mathcal{W}^f$ is strictly increasing in $m$. The surprising feature of the right panel of Figure 4 is not that the world welfare rises in $m$ but that for all $m < \tilde{m}$ the world welfare falls short of its level in the BAU scenario. The coalition of size $m < \tilde{m}$ clearly manages to raise its welfare above the BAU level by increasing the climate damage at the expense of the fringe countries. As the latter engage in costly mitigation to keep the increase in total emissions small, they suffer a welfare loss compared to BAU (left panel of Figure 4) which is even larger than the coalition's welfare gain.

[22] We cannot generalize our findings from Example 1 by induction, of course. Yet we have run several other examples, e.g. Example 2 with the parameters $a = 1000$, $b = 2000$, $\bar{x} = 12$, $\alpha = 50,000$, and $n = 100$ (to be considered in the next section). The graphs corresponding to all examples under scrutiny turned out to be qualitatively the same as those in the Figures 3, 4 and 5 which is why we restrict the graphical presentation to Example 1.

[23] In Example 2 specified in the previous footnote we have calculated the rather large values $\tilde{m} = 42.01$ and $(\tilde{m}/n) = 0.4201$. However, the Tables 1 and 2 below also contain examples in which $\tilde{m}$ as well as $\tilde{m}/n$ are relatively small.

[24] We consider as negligible that the functions $\mathcal{E}^f, \mathcal{W}^f, m\mathcal{W}^c + (n-m)\mathcal{W}^f$ and $m\mathcal{E}^c + (n-m)\mathcal{E}^f$ are slightly non-monotone for $m < 2$.

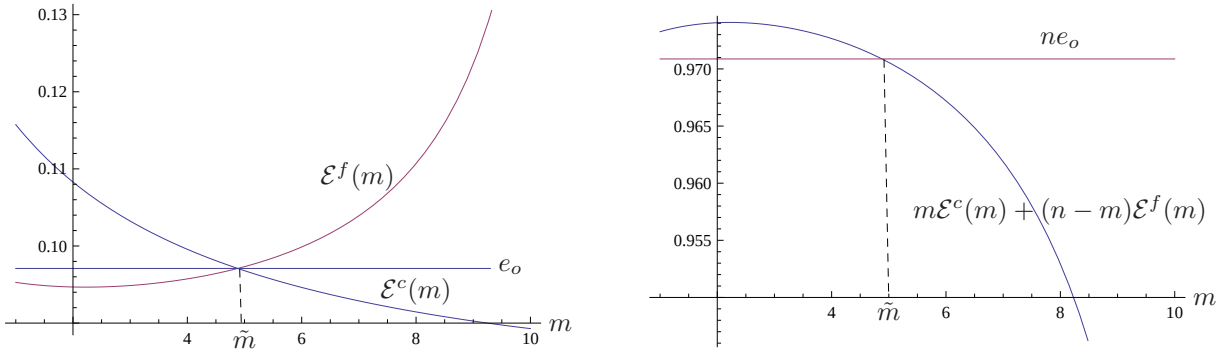


Figure 3: Emissions caps and total emissions in Example 1

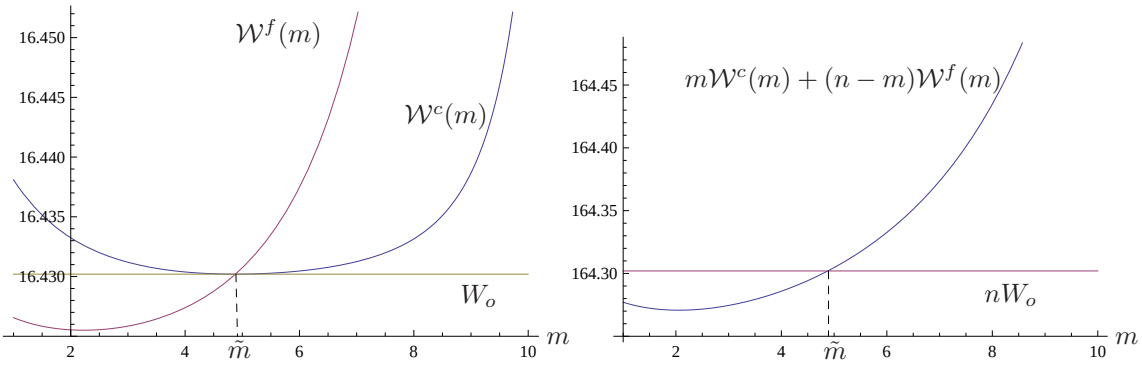


Figure 4: Welfare and aggregate welfare in Example 1

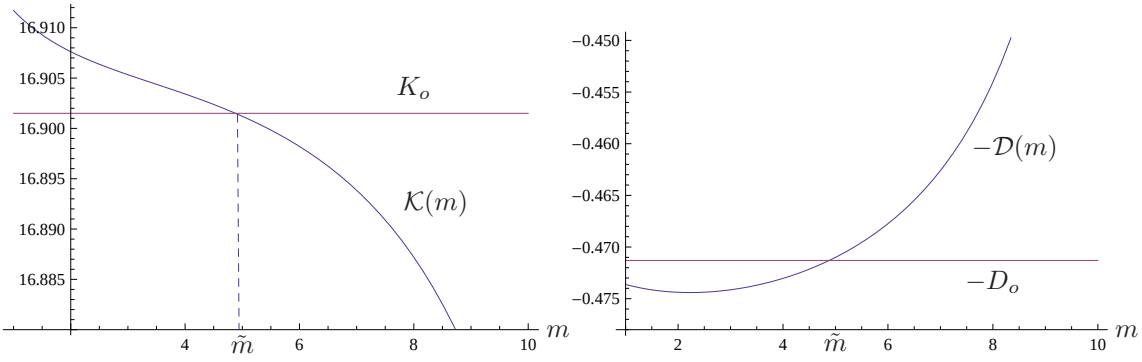


Figure 5: Consumption welfare $[\mathcal{K}(m)]$ and climate welfare $[-\mathcal{D}(m)]$ of the coalition in Example 1

Figure 5 decomposes the (total) welfare of a coalition country into its consumption welfare (curve $\mathcal{K}^c(m)$) and climate welfare (curve $-\mathcal{D}^c(m)$). Figure 5 illustrates that owing to the high leakage rate in case of $m < \tilde{m}$, it is advantageous for the coalition to sacrifice, compared to BAU, some climate welfare for additional consumption welfare. Conversely, if $m > \tilde{m}$, the coalition is more effective in reducing total emissions and therefore benefits from shifting away from consumption welfare toward higher climate welfare.

## 3.2 Self-enforcing IEAs

In the preceding Section 3.1 we have presupposed the presence of a climate coalition and the focus has been on characterizing the Stackelberg equilibrium and its dependence on the exogenous coalition size $m$. Now we turn to the issue of coalition stability. Since supranational authorities for the effective enforcement of agreements are not available, IEAs will not prevail unless they are self-enforcing in the sense that no signatory has an incentive to withdraw (*internal stability*) and no non-signatory has an incentive to sign the agreement (*external stability*).[25] In formal language, an IEA with $m \in \{1, \ldots, n\}$ signatories is said to be self-enforcing or stable if it satisfies the internal stability condition

$$\mathcal{W}^c(m) \geq \mathcal{W}^f(m-1) \tag{32}$$

and the external stability condition

$$\mathcal{W}^f(m) \geq \mathcal{W}^c(m+1). \tag{33}$$

Since the definition of stability requires treating the coalition membership $m$ as an integer, we now have to distinguish between the membership $m \in \{1, \ldots, n\}$ of real-world IEAs and the real-number approximation $m \in [1, n]$ which we have applied before for mathematical convenience. With that distinction in mind we obtain

**Lemma 3.** *If a self-enforcing IEA with $m^* \in \{1, \ldots, n\}$ signatories exists then $m^* > \tilde{m}$.*

To verify Lemma 3 observe that $\mathcal{W}^c(m) > W_o > \mathcal{W}^f(m)$ for all $m < \tilde{m}$ from (30) implies $\mathcal{W}^f(m) < \mathcal{W}^c(m+1)$ and hence the external stability condition is violated for all $m \in \{1, \ldots, n\}$ with $m < \tilde{m}$. If $\tilde{m}$ happens to be an integer, the coalition of size $\tilde{m}$ is not stable either because fringe countries have still an incentive to join the coalition ($W^f(\tilde{m}) < W^c(\tilde{m}+1)$).

The important message of Lemma 3 is that all ('anti-climate') coalitions pushing up total emissions above BAU level fail to be stable. The downside is that Lemma 3 leaves open whether $m^*$ exists and if so how large the positive difference $(m^* - \tilde{m})$ is. Unfortunately, we have not succeeded in resolving these issues analytically. We therefore resort to examining the stability conditions (32) and (33) for the numerical Examples 1 and 2 introduced in the previous Section 3.1. The Figures 6 and 7 present the graphs of the functions $\mathcal{W}^c(m) - \mathcal{W}^f(m-1)$ and $\mathcal{W}^f(m) - \mathcal{W}^c(m+1)$ for the Examples 1 and 2 and their right panels exhibit

[25] This notion of self-enforcement or stability was originally introduced by D'Asprement et al. (1983) in the context of cartel formation.

an enlarged detail of the relevant domain. In both cases there is one and only one interval of coalition sizes in which both functions take on non-negative values (thus satisfying (32) and (33)), and this interval contains one and only one integer, $m^* = 5$ in Example 1 and $m^* = 43$ in Example 2. Moreover, in both cases the stable coalition size $m^*$ is the smallest integer greater than $\tilde{m}$ such that between 40% and 50% of all countries are members of the stable coalition. That contrasts sharply with the result of Rubio and Ulph (2006) and Diamantoudi and Sartzetakis (2006) according to whom the number of signatories in self-enforcing IEAs is small in the parameter sub-space securing positive equilibrium emissions.[26]

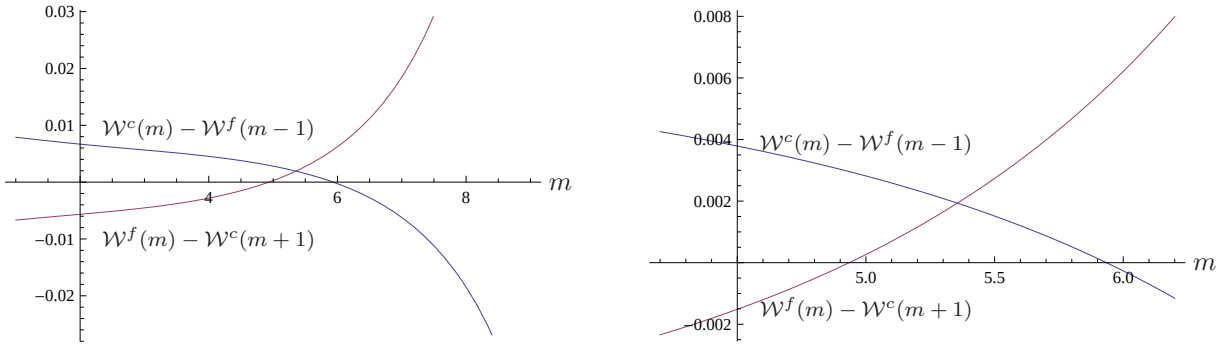


Figure 6: Stability in Example 1 ($\tilde{m} = 4.881, m^* = 5$)

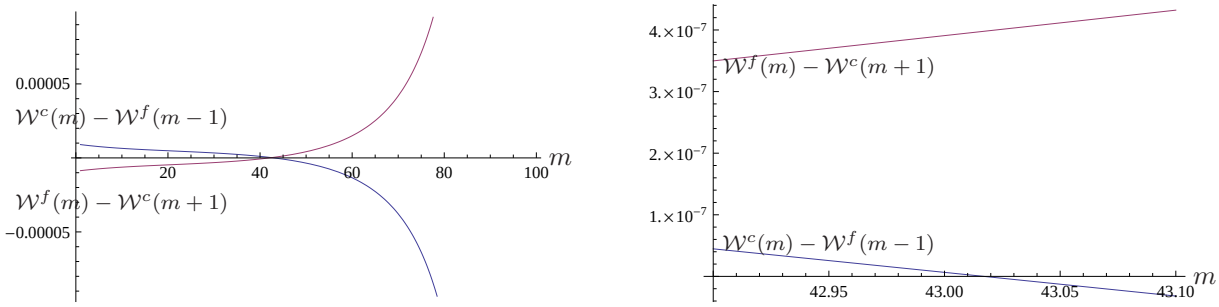


Figure 7: Stability in Example 2 ($\tilde{m} = 42.013, m^* = 43$)

We carried out a number of numerical calculations in addition to the Examples 1 and 2 and their modifications in the Tables 1 and 2 below and reached the unequivocal result that for every parameter constellation securing positive equilibrium emissions there exists a unique self-enforcing IEA whose coalition size $m^*$ is the smallest integer larger than $\tilde{m}$ from (26). Thus it is clear from our comments on Proposition 1 that the Stackelberg equilibrium allocation associated to those self-enforcing IEAs is approximately the same as in BAU; the

[26]It is straightforward from the left panels of Figures 6 and 7 that the equilibrium emissions $\mathcal{E}^f(m^*)$ and $\mathcal{E}^c(m^*)$ are strictly positive.

climate damage is only slightly lower and the coalition countries' welfare is only slightly higher than in BAU while the welfare gain of the free riding fringe countries is greater than that of the coalition countries. The divergence between the Stackelberg equilibrium for $m^*$ and the BAU equilibrium declines in relative terms with an increasing total number, $n$, of countries, and an increasing level of $\tilde{m}$, because the relative impact of changing the coalition size by a fraction of one is declining in $n$ and $\tilde{m}$.

Since $(m^*-\tilde{m}) \leq 1$, we can assess the determinants of the size of $m^*$ by investigating the determinants of $\tilde{m}$. Recall that according to (26), $\tilde{m}$ depends on the size of the parameters $\alpha, b$ and $n$. To examine how $\tilde{m}$ varies with $\alpha$ we differentiate (26) with respect to $\alpha$ and obtain

$$\frac{\mathrm{d}\tilde{m}}{\mathrm{d}\alpha} = \frac{n^2(n-1)[b(n-1)-n]}{[\alpha(2n-1)+n^2(b+1)]^2} \gtreqless 0 \quad \Longleftrightarrow \quad b \gtreqless \frac{n}{n-1}. \tag{34}$$

For $\alpha$ converging to infinity we find $\lim_{\alpha\to\infty} \tilde{m} = \frac{n^2}{2n-1}$.

| $\alpha$ | 1 | 10 | 50 | 100 | 500 | 1000 | $\infty$ |
|---|---|---|---|---|---|---|---|
| $\tilde{m}$ | 1.46 | 1.75 | 2.62 | 3.25 | 4.57 | 4.88 | 5.26 |
| $m^*$ | 2 | 2 | 3 | 4 | 5 | 5 | 6 |

Table 1: Variations of $\alpha$ in Example 1 ($n = 10$)

| $\alpha$ | 1 | $10^3$ | $10^4$ | $10^5$ | $5 \cdot 10^5$ | $10^6$ | $10^7$ | $\infty$ |
|---|---|---|---|---|---|---|---|---|
| $\tilde{m}$ | 1.05 | 1.53 | 5.50 | 25.58 | 42.01 | 45.75 | 49.76 | 50.25 |
| $m^*$ | 2 | 2 | 6 | 26 | 43 | 46 | 50 | 51 |

Table 2: Variations of $\alpha$ in Example 2 ($n = 100$)

According to (34) the comparative static effect of $\alpha$ depends on the size of $b$. The values of $b$ and $n$ chosen in the Examples 1 and 2 satisfy $b > n/(n-1)$ such that $\tilde{m}$ is increasing in $\alpha$ and converges toward $n^2/(2n-1)$ from below. This is confirmed by the numerical examples listed in the Tables 1 and 2. Suppose next that $b < n/(n-1)$. In that case $\tilde{m}$ is decreasing in $\alpha$ and converges toward $n^2/(2n-1)$ from above. That is, for $b < n/(2n-1)$ and $\alpha$ sufficiently small, equation (26) allows for very high levels of $\tilde{m}$, even for $\tilde{m} = n$ (grand coalition). However, the non-negativity constraint for emissions turns out to be violated for low values of $\alpha$ (and $b < n/(n-1)$). We have not succeeded to generate numerical examples of Stackelberg equilibria exhibiting both non-negative emissions and stable coalition sizes larger than $\frac{n^2}{2n-1}$. Hence under the condition of positive equilibrium emissions the maximum share of countries joining a stable coalition, $100\, m^*/n$, appears to be slightly higher than 50%. We also need to emphasize, however, that there are various examples in the Tables 1 and 2 in which the share $100\, m^*/n$ is much smaller than 50%.

The role $\alpha$ plays for the formation of self-enforcing IEAs calls for an economic interpretation. To keep focussed we restrict our attention to the set of parameters satisfying $b > n/(n-1)$ and define the fuel extraction costs, expressed in units of good $X$, as

$$C(e_i^s) := T(0) - T(e_i^s) = \frac{\alpha}{2}(e_i^s)^2 \tag{35}$$

Those extraction costs are obviously progressively increasing such that increasing $\alpha$ corresponds to rising marginal extraction costs. We conclude that an increase in $\alpha$ means that the extraction of fuel becomes more expensive, which increases the size of stable coalitions in turn. The lower and the less progressive the extraction costs are, the smaller is the size of the stable coalition. We summarize our results in

**Proposition 3.** *Under the conditon of positive equilibrium emissions there exist self-enforcing IEAs that are characterized as follows:*

*(i) If $b > n/(n-1)$, then the stable coalition size $m^*$ increases in the parameter $\alpha$ such that as many as (slightly more than) 50% of all countries can be members of a stable coalition.*

*(ii) The number of countries in the self-enforcing IEAs is the smallest integer $m^*$ larger than $\tilde{m}$ from (26) implying that the corresponding Stackelberg equilibrium allocation differs only slightly from the allocation in the scenario of global non-cooperation (BAU).*

We are aware of the limited scope of Proposition 3 because that proposition is based on numerical examples. Nonetheless, the unequivocal result of the calculations we conducted suggests that the messages of Proposition 3 are more general. Proposition 3(i) gives support to the expectation that international trade may lead to rather large stable coalitions. That appears to be good news for proponents of strong climate damage mitigation action if large stable coalitions promise to bring about reductions of global emissions that are larger by an order of magnitude than in BAU and hopefully not too far away from the socially optimal allocation. Unfortunately, Proposition 3(ii) shatters that expectation. Our numerical calculations rather suggest that *all* stable coalitions reduce world emissions only insignificantly compared to BAU emissions. If that result is general - which we are not able to prove analytically - the highly inconvenient implication would be that any attempt to reach a self-enforcing IEA is futile even if the difficult negotiation process would be costless.

Proposition 3(ii) calls for an explanation and the economic intuition. It is clear from the definition of external and internal stability that the number of stable coalitions and their membership depend on the properties the functions $\mathcal{W}^c$ and $\mathcal{W}^f$. We know from our above

analysis of the parametric model and the numerical calculations that

$$\left.\begin{array}{l} \mathcal{W}^f(\tilde{m}) = \mathcal{W}^c(\tilde{m}) = W_o \text{ and } \mathcal{W}^f(m) > \mathcal{W}^c(m) > W_o,\ m > \tilde{m} \text{ (equivalence (33))}, \\ \mathcal{W}^f_m(\tilde{m}) > 0, \mathcal{W}^c_m(\tilde{m}) = 0 \text{ and } \mathcal{W}^h_m(m) > 0, h = c, f, m > \tilde{m} \text{ (Figure 4)}, \\ \mathcal{W}^f_m(m) - \mathcal{W}^c_m(m) > 0 \text{ for } m > \tilde{m} \text{ and increasing in } m \text{ (Figure 10, Appendix D)}, \end{array}\right\} \quad (36)$$

This is important information but unfortunately in the model with the parametric functions (3) the complexity of $\mathcal{W}^c$ and $\mathcal{W}^f$ does not to allow for an analytical characterization of the outcome of stable coalitions. To get an idea, nonetheless, of the relation between $\mathcal{W}^c$, $\mathcal{W}^f$ and $m^*$, define the functions $\Omega^h : [\tilde{m}, n] \longrightarrow \mathbb{R}_+, h = c, f$, by

$$\Omega^c(m) = \omega_o + \frac{\omega_1}{2}(m - \tilde{m})^2 \quad \text{and} \quad \Omega^f(m) = \omega_o + \omega_2(m - \tilde{m}) + \frac{(\omega_1 + \omega_3)}{2}(m - \tilde{m})^2, \quad (37)$$

where the parameters $\omega_o, \omega_1, \omega_2$ and $\omega_3$ are assumed to satisfy $\omega_o = \mathcal{W}^f(\tilde{m}) = \mathcal{W}^c(\tilde{m})$, $\omega_1 > 0$, $\omega_2 = \mathcal{W}^f_m(\tilde{m}) > 0$ and $\omega_3 > 0$. By construction, the functions $\Omega^c$ and $\Omega^f$ satisfy (36) and therefore approximate the functions $\mathcal{W}^c$ and $\mathcal{W}^f$ on the sub-domain $[\tilde{m}, n]$. Taking advantage of that approximation we establish (and prove in the Appendix E)

**Proposition 4.** *Replace the functions $\mathcal{W}^c$ and $\mathcal{W}^f$ on the sub-domain $[\tilde{m}, n]$ by the functions $\Omega^c$ and $\Omega^f$ defined in (37).*

*(i) There exists a stable coalition of size $m^* > \tilde{m}$ and $m^*$ is unique, in general.*[27]

*(ii) The stable coalition size $m^*$ is increasing in $\omega_1$ and declining in $\omega_2$ and $\omega_3$.*

*(iii) If the stable coalition size $m^*$ is the smallest integer greater than $\tilde{m}$, the parameters $\omega_1, \omega_2$ and $\omega_3$ satisfy $3\omega_1 < 2\omega_2 + \omega_3$.*

Proposition 4(i) is in line with our numerical results and supports the view that uniqueness is a rather general result. For an interpretation of Proposition 4(ii), consider the term

$$\frac{\Omega^f_m - \Omega^c_m}{\Omega^c_m} = \frac{\omega_2 + \omega_3 m}{\omega_1 m} \quad (38)$$

which represents the (remaining) fringe countries' relative free-rider benefit from a marginal increase in the coalition size. This benefit is decreasing in $\omega_1$ and increasing in $\omega_2$ and $\omega_3$. Combining that observation with Proposition 4(ii) we find that the stable coalition size $m^*$ is the smaller, the larger is the fringe countries' relative free-rider benefit (38). In other words, the larger is that free-rider benefit, the greater is the coalition countries' incentive to leave the coalition. Proposition 4(iii) confirms that result. We consider $3\omega_1 < 2\omega_2 + \omega_3$ in (38) and conclude that if the stable coalition size $m^*$ is the smallest integer greater than $\tilde{m}$, then the relative free-rider benefit (38) must exceed the level $\frac{3\omega_2 + 3\omega_3 m}{(2\omega_2 + \omega_3)m}$.

[27] We show in the proof (Appendix E) that in exceptional cases there are two stable coalitions.

# 4 On the role of international trade

Up to now we have dealt with a world economy characterized by the four parameters $(a, \alpha, b, n) \in \mathbb{R}^4_{++}$ in the regime of free trade. The straightforward way of improving our understanding of the role of international trade for the formation of self-enforcing IEAs is to compare the results we have derived in the free-trade model with those of the autarky scenario in the otherwise unchanged model. The only substantive modification of the model (1) - (9) is to replace (4) by

$$x_i^s = x_i^d \quad \text{and} \quad e_i^s = e_i^d \qquad i = 1, \ldots, n, \tag{39}$$

which simply turns the world markets for good $X$ and fossil fuel into domestic markets. Good $X$ can still be taken as numéraire ($p_{xi} = 1$ for $i = 1, \ldots, n$) but (5) is now replaced by $p_{ei} = -T'(e_i)$ for $i = 1, \ldots, n$. With these changes the welfare of country $i$ is given by

$$W^i(e_1, \ldots, e_n) = V(e_i) + T(e_i) - D\left(\sum_j e_j\right) \tag{40}$$

for the general functions (1) and (2) and by

$$W^i(e_1, \ldots, e_n) = ae_i - \frac{\check{b}}{2}e_i^2 + \bar{x} - \frac{1}{2}\left(\sum_j e_j\right)^2 \tag{41}$$

with $\check{b} := b + \alpha$ for the parametric functions (3).

The comparison of (12) and (40) subject to (3) shows that the switch from free trade to autarky turns the economy $(a, \alpha, b, n) \in \mathbb{R}^4_{++}$ into the economy $(a, \check{\alpha} = 0, \check{b} = b + \alpha, n)$. The latter obviously has the structure of the basic model of the coalition formation literature in which production and international trade is not modeled.[28] Thus our free-trade versus autarky comparison is also a comparison between the basic model and our trade model. In the following we carry out that comparison in several steps.

To begin with, the BAU equilibria of the economy $(a, \alpha, b, n)$ with and without trade are determined by (15) and hence coincide, because comparative advantage is absent when identical countries are treated equally. Moreover, along the lines of the proof of $\tilde{m}$ in (26) one can show that the coalition size[29]

$$\tilde{m}_a := \frac{\check{b} + n}{\check{b} + 1} \tag{42}$$

[28] See e.g. Finus (2001, equation (3.1)). Diamantoudi and Sartzetakis (2006, equation (1)) as well as Rubio and Ulph (2006, equation (1)) restrict their analysis to the parametric version (41) of the basic model.

[29] In the sequel the autarky regime is indicated by the super- or subscript $a$.

for which the Stackelberg equilibrium (in case of real-number coalitions) is equal to the BAU equilibrium in the economy $(a, \alpha, b, n)$. The comparison of (42) with (26) readily yields $\tilde{m}_a < \tilde{m}$.

In order to understand why $\tilde{m}_a$ is smaller than $\tilde{m}$, consider the economy $(a, \hat{\alpha}, \check{b} = b + \alpha, n)$ with $\hat{\alpha} > 0$ in the regime of free trade and denote - in order to avoid confusion - by $\mu$ the coalition size for which the Stackelberg equilibrium (in case of real-number coalitions) equals the BAU equilibrium in that economy. Under the reasonable condition $\check{b} > n/(n-1)$ discussed above the coalition size $\mu$ declines according to (34), if $\alpha$ is successively reduced.[30] For $\hat{\alpha} = 0$, $\mu$ takes on its minimum $\mu = \tilde{m}_a$ because the economy $(a, \alpha, b, n)$ in autarky coincides with the economy $(a, \hat{\alpha} = 0, \check{b} = b + \alpha, n)$. In case of $\hat{\alpha} > 0$ the economy $(a, \hat{\alpha}, \check{b} = b + \alpha, n)$ is a free-trade world economy with production and with fuel extraction costs $C(e_i^s) := (\hat{\alpha}/2)(e_i^s)^2$ according to (35). Successive reductions of $\hat{\alpha}$ lower these costs and their progressivity until the extraction becomes costless at $\hat{\alpha} = 0$. When fuel is a free good, there is no need and no role for international trade anymore such that the outcomes are the same under open and closed borders. Thus we have demonstrated that we can interpret the economy $(a, \alpha, b, n) \in \mathbb{R}^4_{++}$ in the regime of autarky - as well as the basic model of the literature - as the 'polar case' of a free-trade economy with zero fuel extraction costs ($\hat{\alpha} = 0$). In that perspective the absence of extraction costs is the reason for $\tilde{m}_a < \tilde{m}$.

In the economy $(a, \alpha, b, n) \in \mathbb{R}^4_{++}$ in autarky the fringe countries' best-reply function is characterized by the first-order condition $V'(e_f) + T'(e_f) - D'[me_c + (n-m)e_f] = 0$ which implicitly determines the aggregate best-reply function of the fringe, denoted $s_f = R^a(s_c, m)$. It is straightforward to show that the function $R^a$ exhibits the same qualitative properties as the function $R$ from (16) such that Lemma 1, Lemma 2, (21) - (25) and (27) carry over to the autarky regime. Likewise, Proposition 1 and Lemma 3 still hold when we replace $\tilde{m}$ by $\tilde{m}_a$. An important quantitative difference between both regimes proved in the Appendix F is $\left|R_{s_c}\right| > \left|R^a_{s_c}\right|$, that is, the leakage rate is larger in the free-trade regime than in autarky.[31] As an immediate consequence of (41) the marginal (aggregate) welfare of coalition countries evaluated at BAU (defined in Appendix C) is lower under free trade than under autarky, formally $MWC_o(m) < MWC^a_o(m)$. Since $\tilde{m}$ and $\tilde{m}_a$ are determined by $MWC_o(\tilde{m}) = 0$ and $MWC^a_o(\tilde{m}_a) = 0$, respectively, we infer from (27) and its analogue for autarky that $\tilde{m} > \tilde{m}_a$. Thus we identify $\left|R_{s_c}\right| > \left|R^a_{s_c}\right|$ as a driver for $\tilde{m} > \tilde{m}_a$.

[30] The examples of the Tables 1 and 2 show that $\mu$ may even drop below $\mu = 2$ for small but still positive $\hat{\alpha}$.

[31] Emissions of the fringe and of the coalition are strategic substitutes under both free trade and autarky, but they are stronger strategic substitutes with trade than without. Copeland and Taylor (2005) reach the opposite conclusion in a model that differs substantially from ours - and even find conditions under which emissions of different countries turn into strategic complements when the borders are opened.

To further characterize the differences between autarky and free trade we consider Example 1 for autarky. The 'autarky functions' $\mathcal{E}^{ca}, \mathcal{E}^{fa}, \mathcal{W}^{ca}$ and $\mathcal{W}^{fa}$ turn out to exhibit the same qualitative properties (sign of slope, curvature) as the functions $\mathcal{E}^c, \mathcal{E}^f, \mathcal{W}^c$ and $\mathcal{W}^f$ in Example 1 with free trade. More specifically, the equivalences (28) - (31) (and hence Proposition 2 and Lemma 3) carry over to the autarky scenario when the superscript $a$ is attached to $\mathcal{E}^c, \mathcal{E}^f, \mathcal{W}^c$ and $\mathcal{W}^f$ and $\tilde{m}$ is replaced by $\tilde{m}_a$. In Example 1 the benchmark coalition size in autarky, $\tilde{m}_a = 1.009$, is significantly smaller than its free-trade counterpart $\tilde{m} = 4.881$ confirming our above result on the sign of the difference $\tilde{m} - \tilde{m}_a$. The graphical presentation of these results is delegated to the Appendix G because careful scaling and the plot of enlarged details of the curves are necessary to demonstrate that the analogue of (28) and (30) holds in the case of autarky. In the Figures 8 and 9 below we rather use Example 1 for illustrating what the differences in outcome for the coalition countries are under free trade and autarky and how these differences depend on the coalition size.

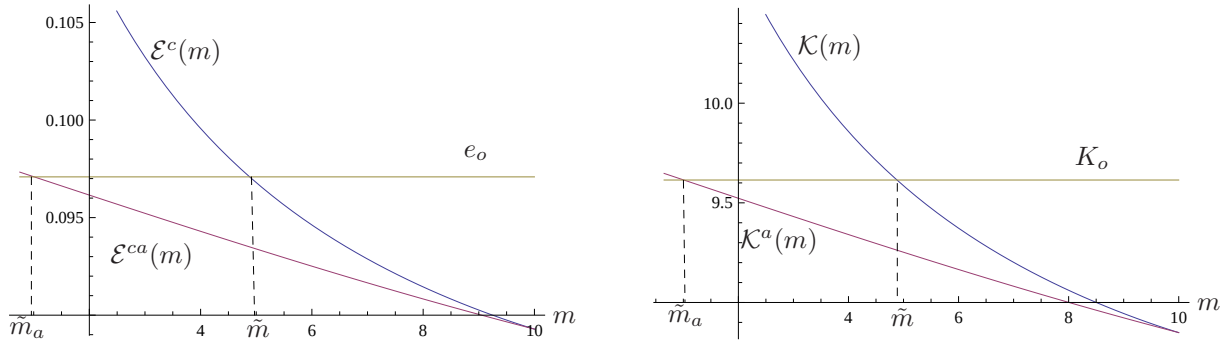


Figure 8: Autarky vs. free trade. Emissions and consumption welfare of coalition countries in Example 1.

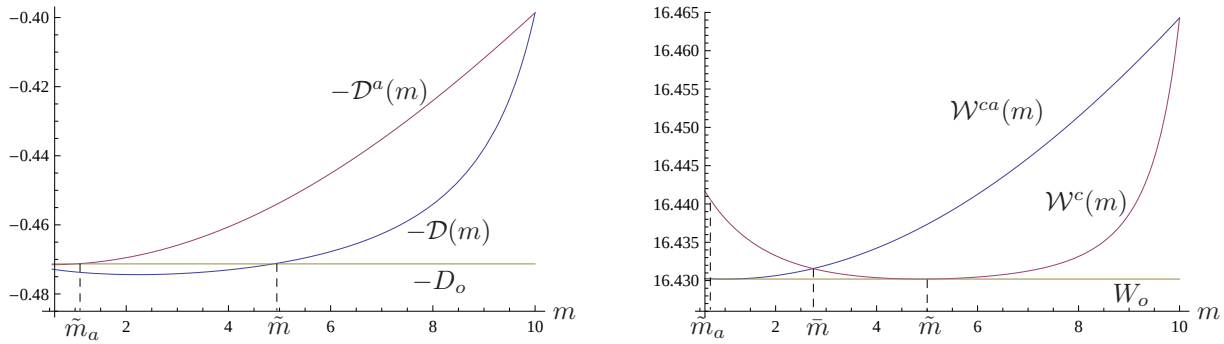


Figure 9: Autarky vs. free trade. Climate damage and total welfare of coalition countries in Example 1.

We restrict our focus on $m \geq \tilde{m}_a$ which is the relevant sub-domain because if a stable coalition exists in autarky, its size is larger than $\tilde{m}_a$ according to the suitably adjusted Lemma 3. The left panel of Figure 8 shows that in both regimes the emissions of coalition

countries are falling, that they are lower in autarky than in the trade regime, and that this difference tends to zero with $m$ approaching $n$. The positive difference $\mathcal{E}^c(m) - \mathcal{E}^{ca}(m)$ is clearly due to $\left|R_{s_c}\right| > \left|R^a_{s_c}\right|$. Moreover, in autarky all countries necessarily consume what they produce, i.e. they choose their consumption as a point on the transformation curve such that $x^d_i = x^s_i = T[\mathcal{E}^{ca}(m)]$. Taking the BAU consumption $[T(e_o), e_o]$ and the corresponding consumption welfare, $K_o$, as a benchmark yields: $T[\mathcal{E}^{ca}(m)] \gtreqless T(e_o) \iff \mathcal{K}^a(m) \lesseqgtr K_o$, where $\mathcal{K}^a(m)$ denotes the level of the coalition countries' equilibrium consumption welfare. The right panel of Figure 8 illustrates that relationship, and it also depicts the coalition countries' equilibrium consumption welfare, $\mathcal{K}(m)$, in the free-trade regime. The latter is larger than $\mathcal{K}^a(m)$, because the fuel consumption $\mathcal{E}^c(m)$ is larger than $\mathcal{E}^{ca}(m)$ and because the coalition countries benefit from the possibility to decouple consumption from production. The consequence of the more stringent emission reduction in autarky is that the climate welfare of all countries is higher in autarky than in free trade - as illustrated in the left panel of Figure 9.

To sum up, for every $m \in [\tilde{m}_a, n]$ the coalition countries' consumption welfare is lower and their climate welfare is higher in autarky than in free trade implying that their net welfare change is ambiguous. More specific information on the comparison of total welfare $\mathcal{W}^{ca}(m)$ and $\mathcal{W}^c(m)$ provides the right panel of Figure 9. Since the graphs of $\mathcal{W}^{ca}$ and $\mathcal{W}^c$ attain their minimum at $m = \tilde{m}_a$ and $m = \tilde{m}$, respectively, there must be $\bar{m} \in ]\tilde{m}_a, \tilde{m}[$, defined by $\mathcal{W}^{ca}(\bar{m}) = \mathcal{W}^c(\bar{m})$, such that $\mathcal{W}^{ca}(m) < \mathcal{W}^c(m)$ for $m < \bar{m}$ and $\mathcal{W}^{ca}(m) > \mathcal{W}^c(m)$ for all $m$ in some interval $]\bar{m}, \mathring{m}[$ with[32] $\mathring{m} > \tilde{m}$. In other words, when moving from autarky to free trade, the coalition countries' climate welfare gain is overcompensated by their consumption welfare loss and the opposite holds for relatively large coalition sizes $m \in ]\bar{m}, \mathring{m}[$.

As shown above, in the regime of autarky the model of the present paper coincides with the basic model of the coalition formation literature. As a consequence, we can invoke the results of Diamantoudi and Sartzetakis (2006) and Rubio and Ulph (2006) who show that "... restricting parameter values to guarantee interior solutions is a sufficient condition to get stable IEAs with a small number of signatories ..." (Rubio and Ulph, 2006, p. 236). Diamantoudi and Sartzetakis focus exclusively, as we do, on subsets of parameters leading to positive equilibrium emissions and find that stable IEAs have at most four signatories even if the total number of countries is large. Rubio and Ulph (2006) consider a larger parameter space and introduce, in contrast to Barrett (1994), non-negativity constraints on emissions. For a subset of parameter values which guarantee interior solutions they find that the maximum stable coalition size is three.

[32] In the right panel of Figure 9 we find $\mathring{m} = n$, but it is not clear whether that is a general feature.

Barrett (1994) shows that there are parameter constellations for which the self-enforcing IEA may attain any size from very small to the grand coalition. That finding seems to be at variance with the results reported in the last paragraph, but they are not inconsistent for the following reasons. Barrett takes abatement efforts as the governments' choice variable rather than emission caps and does not rule out negative emissions. Diamantoudi and Sartzetakis (2006) convert Barrett's approach into the basic model of type (41) and show that in Barrett's framework self-enforcing IEAs consist of no more than four countries on the set of parameters guaranteeing positive equilibrium emissions.

We conclude that as long as solutions with non-positive emissions are ruled out we get stable IEAs with a small number of signatories in the autarky scenario (= basic model) irrespective of the total number of countries. That result clearly is in stark contrast to our finding in the free-trade model of Section 3 where we have identified stable coalitions much larger than in the autarky model.

Regarding the comparison of free trade and autarky we also want to know how effective the stable coalition is in reducing world emissions below BAU emissions. Rubio and Ulph (2006) do not address that issue. Diamontoudi and Sartzetakis (2006) find that the welfare of the signatories is very close to its lowest value when the IEA is stable but they do not link that observation to the BAU scenario. We will establish that link in several steps. Proposition 1 and Lemma 3 applied to autarky now reads: If a self-enforcing IEA with $m_a^* \in \{1, \ldots, n\}$ exists then $m_a^* \geq \tilde{m}_a$. We restrict our subsequent analysis to the parameters space $\Lambda := \{(\check{b}, n) \,|\, \check{b} > n(n-4)/4, n > 4\}$ and invoke the results of Rubio and Ulph (2006) that for all $(\check{b}, n) \in \Lambda$ the equilibrium emissions are positive (ibidem, footnote 16) and $m_a^* \leq 3$ (ibidem, Corollary 2).

In order to obtain further information about the size of the positive difference $m_a^* - \tilde{m}_a$, we insert $\check{b} = n(n-4)/4$ in (42) and make use of $\frac{\mathrm{d}\tilde{m}_a}{\mathrm{d}b} < 0$ to obtain[33]

$$\tilde{m}_a \in ]1, \bar{M}^a(n)[ \quad \text{where } \bar{M}^a(n) := \frac{n^2}{n^2 - 4(n-1)}. \tag{43}$$

Closer inspection of (43) reveals that $\bar{M}^a(5) = 2.77$ and that $\frac{\mathrm{d}\bar{M}^a(n)}{\mathrm{d}n} < 0$ for $n > 4$. Hence we get

$$\tilde{m}_a \in ]1, 2.77[ \quad \text{for all } (\check{b}, n) \in \Lambda. \tag{44}$$

In view of (42) and (44) and $m_a^* \leq 3$ we conclude that $m_a^* - \tilde{m}_a \leq 2$ for all $n > 4$. Our numerical simulations show (as in case of international trade) that in autarky models with parameters satisfying $(\check{b}, n) \in \Lambda$ the size $m_a^*$ of self-enforcing IEAs is the smallest integer

[33] It follows directly from (42) that 1 is a lower bound for $\tilde{m}_a$.

larger than $\tilde{m}_a$ which reconfirms Corollary 2 of Rubio and Ulph (2006). We summarize these findings in

**Proposition 5.** *Consider the world economy without international trade for the parameter space* $\Lambda := \{(\check{b}, n) \big| \check{b} > \frac{n(n-4)}{4}, n > 4\}$.

*(i) Then our model coincides with the models of Diamantoudi and Sartzetakis (2006) and Rubio and Ulph (2006).*

*(ii) Then the size* $m_a^*$ *of self-enforcing IEAs is the smallest integer larger than* $\tilde{m}_a$ *from (42), and at most equal to 3.*

*(iii) The emission caps implemented by the self-enforcing IEA are only slightly tighter than the emission cap in the BAU equilibrium under autarky.*

# 5 Concluding remarks

The present paper reexamines the issue of self-enforcing international environmental agreements (IEAs) extending the basic model of the IEA literature introduced by Barrett (1994) and others to a general equilibrium framework with production, consumption and international trade. In models yielding positive equilibrium emissions and with an IEA acting as Stackelberg leader we show

(a) that in stark contrast to the outcome of the basic model large stable IEAs may form,

(b) and that in all Stackelberg equilibria with a stable IEA the 'gains of cooperation' are negligible: Compared to the case of global non-cooperation the coalition countries' welfare gain as well as the climate damage reduction are very small.

While result (a) raises hopes for successful and effective cooperation in fighting climate change, result (b) thwarts these hopes because efforts of achieving effective mitigation through forming a self-enforcing IEA are futile irrespective of how large these IEAs are.

The only major implication specific to modeling international trade turns out to be the finding that under certain conditions stable IEAs with a large number of signatories emerge. But that distinctive feature is inconsequential because neither small nor large self-enforcing IEAs bring about substantial gains of cooperation (result (b)). An interesting side result is that in the absence of international trade our model of autarkic countries coincides with the basic model of the extant IEA literature. That is, the basic model can be interpreted as a model of autarkic countries. We infer from the extant literature that in this autarky scenario the number of signatories of the self-enforcing IEA is very small, and we show that

the allocation in the corresponding Stackelberg equilibrium does not differ much from the business-as-usual allocation. As in the case of international trade, the coalition countries' welfare rises and the climate damage declines by a very small amount only.

Although our model has more 'economic' structure than the basic model we have kept it simple enough for the benefit of comparing it with the basic model and for the benefit of deriving informative results. As pointed out in the introduction the assumption of emissions being non-essential is not fully satisfactory for carbon emissions in the context of climate change mitigation. It is necessary and desirable to examine the outcome for the case of essential emissions even if analytical results then cannot be obtained anymore. More generally, one would want to check the robustness of results when economies are modeled in a more complex way, e.g. when fossil fuel is not only a final consumption good but also an intermediary industrial input. It is almost needless to say that while the assumption of symmetric countries is crucial for deriving meaningful (analytical) results, it abstracts from many real-world complexities which are severe barriers to reaching self-enforcing IEAs, and it therefore massively underestimates the difficulties of forming such agreements.

# References


Barrett, S. (1994): Self-enforcing international environmental agreements. *Oxford Economic Papers* 46, 878-894.

Barrett, S. (1997): The strategy of trade sanctions in international environmental agreements. *Resource and Energy Economics* 19, 345-361.

Barrett, S. (1999): A theory of full international cooperation. *Journal of Theoretical Politics* 11, 519-541.

Barrett, S. (2001): International cooperation for sale. *European Economic Review* 45, 1835-1850.

Buchner, B., Carraro, C. and I. Cersosimo (2002): Economic consequences of the US withdrawal from the Kyoto/Bonn Protocol, *Climate Policy* 2, 273-292.

Carraro, C. and D. Siniscalco (1991): Strategies for the international protection of the environment. *CEPR discussion paper* 568.

Carraro, C. and D. Siniscalco (1993): Strategies for the international protection of the environment. *Journal of Public Economics* 52, 309-328.

Carbone, J.C., Helm, C. 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, 233-263.

Copeland, B.R. and M.S. Taylor (2005): Free trade and global warming: a trade theory view of the Kyoto protocol. *Journal of Environmental Economics and Management* 49, 205-234.

D'Aspremont, C., Jacquemin, A, Gabszewicz, J.J. and J.A. Weymark (1983): On the stability of collusive price leadership. *Canadian Journal of Economics* 16, 17-25.

Diamantoudi, E. and E. Sartzetakis (2006): Stable international environmental agreements: An analytical approach. *Journal of Public Economic Theory* 8, 247-263.

Eichner, T. and R. Pethig (2012): Sub-global climate coalition and international trade, *CESifo working paper* 3915.

Finus, M. (2003): Stability and design of international environmental agreements: the case of transboundary pollution, in: H. Folmer and T. Tietenberg (eds.), *The International Yearbook of Environmental and Reource Economics* 2003/2004, Edward Elgar, Cheltenham, 82-158.

Hannesson, R. (2010): The coalition of the willing: Effect of country diversity in an international treaty game. *Review of International Organizations* 5, 461-474.

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

Hoel, M. (1992): International environmental conventions: the case of uniform reductions of emissions *Environmental and Resource Economics* 2, 141-159.

Hoel, M. and K. Schneider (1997): Incentives to participate in an international environmental agreement. *Environmental and Resource Economics* 9, 153-170.

Kolstad, C. (2007): Systematic uncertainty in self-enforcing international environmental agreements. *Journal of Environmental Economics and Management* 53, 68-78.

Rubio, S.J. and A. Ulph (2006): Self-enforcing agreements and international trade in greenhouse emission rights. *Oxford Economic Papers* 58, 233-263.

# Appendix

## Appendix A: Proof of Lemma 1

(i) Inserting the parametric functions (3) in (13) yields, after rearrangement of terms

$$e_i = \underbrace{\frac{an^2}{\alpha(2n-1)+(1+b)n^2}}_{=:G} - \underbrace{\frac{\alpha(n-1)+n^2}{\alpha(2n-1)+(1+b)n^2}}_{:=H} \sum_{j\neq i} e_j \quad \text{for} \quad i=1,\ldots,n. \tag{A1}$$

From (A1) we get

$$e_i = G - H\left(\sum_{j\in C} e_j + \sum_{j\in F, j\neq i} e_j\right) = G - Hme_C - H\sum_{j\in F, j\neq i} e_j \quad \text{for all}\, i\in F. \tag{A2}$$

Summing over $i \in F$ yields

$$\sum_{i\in F} e_i = (n-m)e_f = (n-m)G - (n-m)Hme_c - (n-m-1)H(n-m)e_f \tag{A3}$$

which can be rearranged to

$$(n-m)e_f = \frac{(n-m)G}{1+(n-m-1)H} - \frac{(n-m)H}{1+(n-m-1)H} me_c. \tag{A4}$$

or equivalently to

$$s_f = R(s_c, m) = \frac{(n-m)G}{1+(n-m-1)H} - \frac{(n-m)H}{1+(n-m-1)H} s_c. \tag{A5}$$

Next, verify that $\hat{s}_c := R^{-1}(s_f = 0, m) = \frac{G}{H}$ is independent of $m$. Finally, differentiation of (A5) yields

$$\begin{aligned} R_m &= -\frac{(1-H)G}{[1+(n-m-1)H]^2} + \frac{(1-H)H}{[1+(n-m-1)H]^2} s_c = -\frac{(1-H)R(s_c,m)}{(n-m)[1+(n-m-1)H]}, \\ R_{s_c} &= -\frac{(n-m)H}{1+(n-m-1)H} < 0, \quad R_{s_c s_c} = 0, \quad R_{s_c m} = \frac{H(1-H)}{[1+(n-m-1)H]^2} > 0 \end{aligned} \tag{A6}$$

due to $G > 0$ and $H \in [0,1]$. ■

## Appendix B: Proof of Lemma 2

Since the coalition size $m$ is constant throughout this proof we omit for convenience $m$ as argument of the welfare functions. We first show the strict concavity of the coalition country's welfare function. Total differentiation of $W^c(s_c, \underbrace{R(s_c)}_{=s_f})$ from (17) yields

$$\frac{\mathrm{d}W^c}{\mathrm{d}s_c} = W^c_{s_c} + W^c_{s_f} R_{s_c}, \tag{B1}$$

$$\frac{\mathrm{d}^2 W^c}{\mathrm{d}s_c^2} = \underbrace{W^c_{s_c s_c} + W^c_{s_c s_f} R_{s_c}}_{\equiv \frac{\mathrm{d}W^c_{s_c}}{\mathrm{d}s_c}} + \underbrace{\left[W^c_{s_f s_c} + W^c_{s_f s_f} R_{s_c}\right]}_{\equiv \frac{\mathrm{d}W^c_{s_f}}{\mathrm{d}s_c}} R_{s_c} + W^s_{s_f} \underbrace{R_{s_c s_c}}_{=0}. \tag{B2}$$

Partial differentiation of

$$W^c_{s_c}(s_c, s_f) = \frac{V'\left(\frac{s_c}{m}\right)}{m} + \frac{T'\left(\frac{s_c+s_f}{n}\right)}{m} - \frac{[ms_f - (n-m)s_c]\,T''\left(\frac{s_c+s_f}{n}\right)}{n^2 m} - D'(s_c + s_f) \quad \text{(B3)}$$

yields

$$\begin{aligned} W^c_{s_c s_c} &= \frac{V''}{m^2} + \frac{(2n-m)T''}{n^2 m} - \frac{[ms_f - (n-m)s_c]\,T'''}{n^3 m} - D'' \\ &= -\frac{b}{m^2} - \frac{\alpha(2n-m)}{n^2 m} - \delta, \quad \text{(B4)} \\ W^c_{s_c s_f} &= \frac{(n-m)T''}{n^2 m} - \frac{[ms_f - (n-m)s_c]\,T'''}{n^3 m} - D'' = -\frac{\alpha(n-m)}{n^2 m} - \delta. \quad \text{(B5)} \end{aligned}$$

Making use (B4), (B5) and $R_{s_c} = -\underbrace{\frac{(n-m)H}{(1-H)+(n-m)H}}_{=:\tilde{H}}$ (which follows from differentiation of (A5)) we get

$$\frac{\mathrm{d}W^c_{s_c}}{\mathrm{d}s_c} = -\frac{b}{m^2} - \frac{\alpha(2n-m)}{n^2 m} - \delta + \left[\frac{\alpha(n-m)}{n^2 m} + \delta\right]\tilde{H}. \quad \text{(B6)}$$

Partial differentiation of

$$W^c_{s_f}(s_c, s_f) = -\frac{[ms_f - (n-m)s_c]\,T''\left(\frac{s_c+s_f}{n}\right)}{n^2 m} - D'(s_c + s_f). \quad \text{(B7)}$$

yields

$$\begin{aligned} W^c_{s_f s_c} &= \frac{(n-m)T''}{n^2 m} - \frac{[ms_f - (n-m)s_c]\,T'''}{n^3 m} - D'' = -\frac{(n-m)\alpha}{n^2 m} - \delta, \quad \text{(B8)} \\ W^c_{s_f s_f} &= -\frac{T''}{n^2} - \frac{[ms_f - (n-m)s_c]\,T'''}{n^3 m} - D'' = \frac{\alpha}{n^2} - \delta. \quad \text{(B9)} \end{aligned}$$

Making use of (B8), (B9) and $R_{s_c} = -\tilde{H}$ we obtain

$$\frac{\mathrm{d}W^c_{s_f}}{\mathrm{d}s_c} = -\frac{(n-m)\alpha}{n^2 m} - \delta - \left(\frac{\alpha}{n^2} - \delta\right)\tilde{H}. \quad \text{(B10)}$$

Finally, inserting (B6) and (B10) in (B2) establishes

$$\begin{aligned} \frac{\mathrm{d}^2 W^c}{\mathrm{d}s_c^2} &= -\frac{b}{m^2} - \frac{\alpha(2n-m)}{n^2 m} - \delta + \left[\frac{\alpha(n-m)}{n^2 m} + \delta\right]2\tilde{H} + \left(\frac{\alpha}{n^2} - \delta\right)\tilde{H}^2 \\ &= -\frac{b}{m^2} - \frac{\alpha(1-\tilde{H})[2n - (1-\tilde{H})m]}{n^2 m} - \delta(1-\tilde{H})^2 \quad \text{(B11)} \end{aligned}$$

which is negative due to $\tilde{H} \in ]0, 1[$.

Next, we prove the monotonicity property of the fringe country's welfare function. Differentiation of $W^f(s_c, \underbrace{R(s_c)}_{=s_f})$ from (18) yields

$$\frac{\mathrm{d}W^f}{\mathrm{d}s_c} = W^f_{s_c} + W^f_{s_f} R_{s_c}, \quad \text{(B12)}$$

where

$$W^f_{s_c} = \frac{[ms_f - (n-m)s_c]\,T''}{n^2(n-m)} - D' \tag{B13}$$

$$W^f_{s_f} = \frac{V'}{n-m} + \frac{T'}{n-m} + \frac{[ms_f - (n-m)s_c]\,T''}{n^2(n-m)} - D'. \tag{B14}$$

Taking advantage of the fringe countries first-order condition (13) which is equivalent to

$$V' + T' + \frac{[ms_f - (n-m)s_c]\,T''}{n^2(n-m)} - D' = 0 \tag{B15}$$

in (B13) and (B14) we obtain

$$W^f_{s_c} = -(V' + T'), \tag{B16}$$

$$W^f_{s_f} = -\frac{n-m-1}{n-m}(V' + T'). \tag{B17}$$

Inserting (B16) and (B17) in (B12) we get

$$\frac{\mathrm{d}W^f}{\mathrm{d}s_c} = -(V' + T')\left[1 + \frac{n-m-1}{n-m}R_{s_c}\right]. \tag{B18}$$

Since the terms in rectangular brackets are positive, it holds $\frac{\mathrm{d}W^f}{\mathrm{d}s_c} < 0$ if and only if $V' + T' > 0$. From (5) and $V'(e_f) = p + \pi_f$ (which follows from the fringe countries' consumers utility maximization) we have $V' + T' = \pi_f$. From (13) we infer that $V' + T' > 0$ if $e_f > e_c$. Finally, it can be shown that $\pi_f$ remains positive when the coalition relaxes its emission cap and the fringe countries tighten their emission caps. ■

## Appendix C: Proof of Proposition 1

Account for $\frac{\mathrm{d}(s_c+s_f)}{\mathrm{d}s_c} = 1 + R_{s_c}$, and determine the first-order condition for an interior solution to (19),

$$\begin{aligned}\frac{\mathrm{d}(mW^c)}{\mathrm{d}s_c} &= W^c_{s_c} + W^c_{s_f}R_{s_c} \\ &= V' + T' - \left(\frac{s_c + s_f}{n} - \frac{s_c}{m}\right)\frac{m(1+R_{s_c})T''}{n} - m(1+R_{s_c})D' = 0.\end{aligned} \tag{C1}$$

If the coalition of any size $m \in [1, n[$ chooses the strategy $s_c = me_o$, the fringe's best reply is $s_f = R(me_o, m) = (n-m)e_o$ and the BAU equilibrium results. At that equilibrium, i.e. evaluated at $s_c = me_o$, the coalition's marginal welfare is

$$\begin{aligned}MWC_o(m) &:= \frac{\mathrm{d}(mW^c)}{\mathrm{d}s_c}\Big|_{s_c = me_o} = \\ &= \underbrace{V'(e_o) + T'(e_o)}_{\text{marginal consumption welfare, same for all coaltion sizes}} + \underbrace{\{-D'(ne_o) + [1 - m(1+R_{s_c}]D'(ne_o)\}}_{\text{marginal climate welfare for } m \in ]1,n[}.\end{aligned} \tag{C2}$$

According to (C2) the coalition's marginal consumption welfare is independent of $m$, while its marginal climate welfare is not. Since by definition of $\tilde{m}$ the condition $\tilde{m}[1+R_{s_c}(\tilde{m}e_o, \tilde{m})] = 1$ is satisfied, the equations (C1) and (C2) yield for a coalition of size

$$MWC_o(\tilde{m}) = \underbrace{V'(e_o) + T'(e_o)}_{\text{marginal consumption welfare}} + \underbrace{[-D'(ne_o)]}_{\text{marginal climate welfare for } m = \tilde{m}} = 0. \tag{C3}$$

(C3) is identical to (15), the first-order condition of all $n$ countries in the non-cooperative BAU scenario of Section 2. We invoke (C3) to rewrite (C2) as

$$\begin{aligned} MWC_o(m) &= \underbrace{V'(e_o) + T'(e_o) - D'(ne_o)}_{=0} + [1 - m(1 + R_{s_c})]D'(ne_o) \\ &= [1 - m(1 + R_{s_c})]D'(ne_o). \end{aligned} \tag{C4}$$

(C4) holds for any given $m \in [1, n[$. Since $\frac{\mathrm{d}[m(1+R_{s_c})]}{\mathrm{d}m} = (1 + R_{s_c}) + mR_{s_cm} > 0$, the equivalence $\left\{[1 - m(1 + R_{s_c})] \gtreqless 0 \iff m \lesseqgtr \tilde{m}\right\}$ holds. Finally, differentiation of (A5) with respect to $s_c$ yields $R_{s_c} = -\frac{(n-m)H}{1+(n-m-1)H}$. Inserting this term in $[1 - \tilde{m}(1 + R_{s_c})] = 0$ we get $\tilde{m} = 1 + (n - 1)H$. Making use of the definition of $H$ from (A1) establishes after rearrangement of terms (26). ■

## Appendix D: $\mathcal{W}^f_m - \mathcal{W}^c_m$ in Example 1

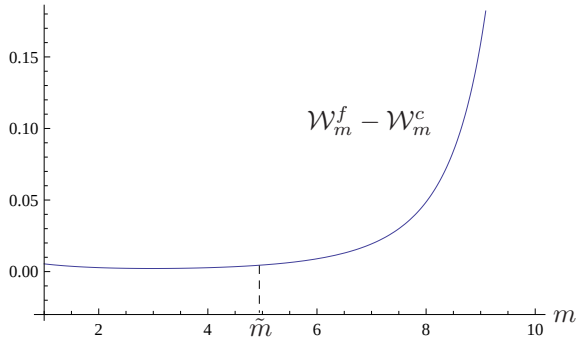


Figure 10: $\mathcal{W}^f_m - \mathcal{W}^c_m$ in Example 1

## Appendix E: Proof of Proposition 4

Ad (i): Define the function $H : ]\tilde{m}, n[ \longrightarrow \mathbb{R}_+$ by $[h = H(m) \iff \mathcal{W}^f(m) = \mathcal{W}^c(m + h)]$.[34] From $\mathcal{W}^f(\tilde{m}) = \mathcal{W}^c(\tilde{m})$ and $\mathcal{W}^f(m) > \mathcal{W}^c(m)$ for all $m > \tilde{m}$ follows $H(\tilde{m}) = 0$ and $H(m) > 0$ for all $m > \tilde{m}$, and hence $H_m(\tilde{m}) > 0$. Let $\tilde{m}^{(+)}$ be the smallest integer greater than $\tilde{m}$ and take advantage of the function $H$ to characterize a stable coalition as follows:

[34] Graphically speaking, if $m$ is plotted on the horizontal axis, $H(m)$ is the horizontal distance between the $\mathcal{W}^f$ curve and the $\mathcal{W}^c$ curve at the level $\mathcal{W}^f(m)$ above the $m$-axis.

*The coalition of size* $m^* \in \{\tilde{m}^{(+)} \ldots, n\}$ *is stable, if and only if there is* $\hat{m} \in ]\tilde{m}, n[$ *such that* $\hat{m} \leq m^*$, $\hat{m}+1 \geq m^*$, $H(\hat{m}) = 1$, $B(m^*) \geq 1$ *and* $H(\check{m}) \leq 1$, *where the number* $\check{m}$ *is defined by the equation* $\mathcal{W}^f(\check{m}) = \mathcal{W}^c(m^*)$.

If $\hat{m}$ happens to be an integer, then the coalitions of size $\hat{m}$ and $\hat{m}+1$ are both stable. Otherwise $m^* \in ]\hat{m}, \hat{m}+1[$. Since we cannot determine the shape of the function $H$ based on the parametric functions (3), we approximate it by construction the function $H$ associated with the functions $\Omega^c$ and $\Omega^f$ from (36). We do so by solving the equation $\Omega^f(m) = \Omega^c(m+b)$ and obtain

$$h = H(m, \omega_1, \omega_2, \omega_3) = -(m - \tilde{m}) + \sqrt{\frac{2\omega_2(m-\tilde{m}) + (\omega_1+\omega_3)(m-\tilde{m})^2}{\omega_1}}. \tag{E1}$$

From (E1) we get $H(m) = 0$ for $m = \tilde{m}$ and $H(m) > 0$ for $m > \tilde{m}$. For all $m \geq \tilde{m}$ the first derivative is

$$H_m = -1 + \rho > 0, \quad \text{where } \rho := \sqrt{1 + \frac{\omega_2^2 + [2\omega_2 + (\omega_1+\omega_3)(m-\tilde{m})]\omega_3(m-\tilde{m})}{[2\omega_2 + (\omega_1+\omega_3)(m-\tilde{m})]\omega_1(m-\tilde{m})}}. \tag{E2}$$

$H(0) = 0$ and (E2) imply that there is one and only one $\hat{m} \in ]\tilde{m}, n[$ satisfying $H(\hat{m}) = 1$. Hence if $\hat{m}$ is an integer and $\hat{m} > \tilde{m}$, the coalitions of size $\hat{m}$ and size $\hat{m}+1$ are stable coalitions. Otherwise, there exists one and only one stable coalition. Its size is the (unique) integer in the interval $]\hat{m}, \hat{m}+1[$.

Ad (ii): Verify $H_{\omega_1} = -\frac{2\omega_2(m-\tilde{m})+\omega_3(m-\tilde{m})^2}{2\rho\omega_1^2} < 0$, $H_{\omega_2} = \frac{m-\tilde{m}}{\rho\omega_1} > 0$, $H_{\omega_3} = \frac{(m-\tilde{m})^2}{2\rho\omega_1} > 0$ and observe that the differential of $H(\hat{m}, \omega_1, \omega_2, \omega_2) = 1$ yields $\frac{\partial\hat{m}}{\partial\omega_i} = -\frac{H_{\omega_i}}{H_m}$ for $i = 1, 2, 3$. Therefore sign $\frac{\partial\hat{m}}{\partial\omega_i} = -\text{sign}\, H_{\omega_i}$.

Ad (iii): Insert $m = \hat{m}$ and $H(\hat{m}) = 1$ in (E1) to obtain, after some rearrangement of terms,

$$\hat{m} - \tilde{m} = -\frac{\omega_2 - \omega_1}{\omega_3} + \sqrt{\frac{(\omega_2-\omega_1)^2 + \omega_1\omega_3}{\omega_3^2}} > 0. \tag{E3}$$

Denote by $\tilde{m}^{(+)}$ the smallest integer greater than $\tilde{m}$ and recall that $m^* \in [\hat{m}, \hat{m}+1]$. It follows that $(\hat{m} - \tilde{m}) < 1$ is a necessary condition for $m^* = \tilde{m}^{(+)}$. Invoking (E3), we find that $(\hat{m} - \tilde{m}) < 1$ holds, if and only if $3\omega_1 < 2\omega_2 + \omega_3$. ■

## Appendix F: Proof of $\left|R_{s_c}\right| > \left|R^a_{s_c}\right|$

Making use of the parametric functions in the fringe country's first-order condition $V'(e_i) + T'(e_i) - D'\left(\sum_j e_j\right) = 0$ yields

$$e_i = \frac{\alpha}{\alpha+b+1} - \frac{1}{\alpha+b+1}\sum_{j\neq i} e_j. \tag{F1}$$

Multiplying (F1) by $(n-m)$ and setting $e_i = e_f = \frac{s_f}{n-m}$ and $\sum_{j\neq i} e_j = me_c + (n-m-1)e_f = s_c + \frac{n-m-1}{n-m}s_f$ we obtain after rearrangement of terms the aggregate fringe best reply function

$$s_f = R^a(s_c, m) := \frac{(n-m)\alpha}{\alpha+b+n-m} - \frac{n-m}{\alpha+b+n-m}s_c. \tag{F2}$$

Next, differentiating (A5) and (F2) we get

$$\left|R^a_{s_c}\right| < \left|R^a_{s_c}\right| \iff \frac{1}{\alpha+b+n-m} < \frac{H}{1+(n-m-1)H} \iff \frac{1}{H} < 1+\alpha+b. \tag{F3}$$

Inserting $H$ from (A1) in (F3) and rearranging terms establishes

$$\left|R^a_{s_c}\right| < \left|R^a_{s_c}\right| \iff \alpha n < \alpha(\alpha+b)(n-1) + \alpha n^2. \tag{F4}$$

■

## Appendix G: Example 1 for autarky (only for the referees)

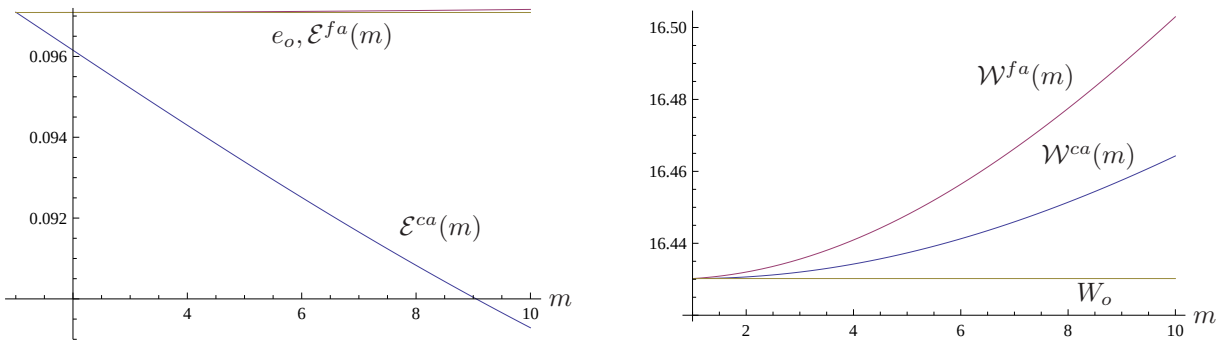


Figure 11: Emissions and welfare in Example 1 for autarky

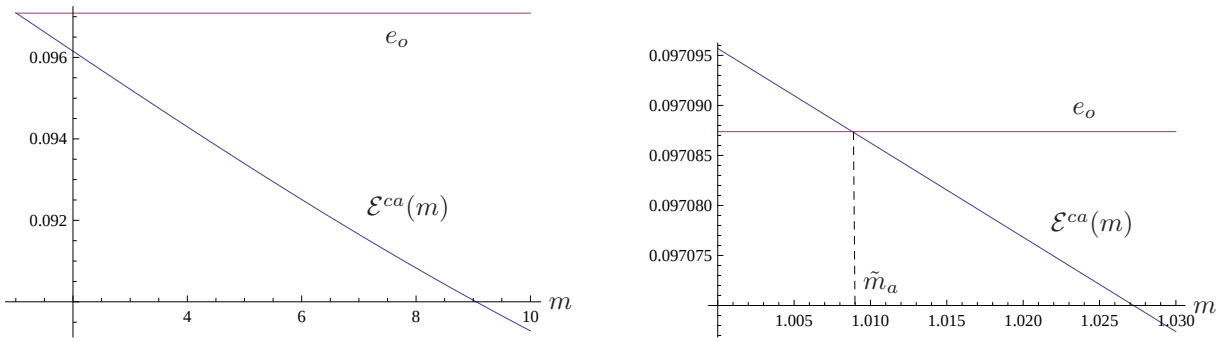


Figure 12: Emissions in Example 1 for autarky

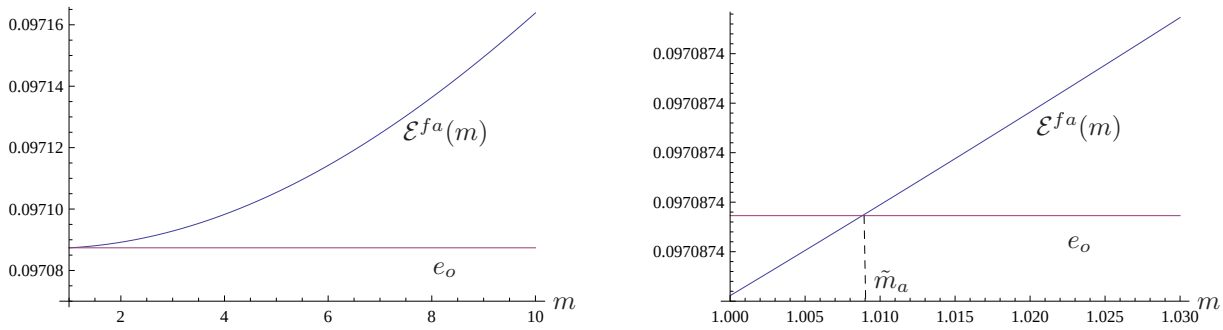


Figure 13: Emissions in Example 1 for autarky

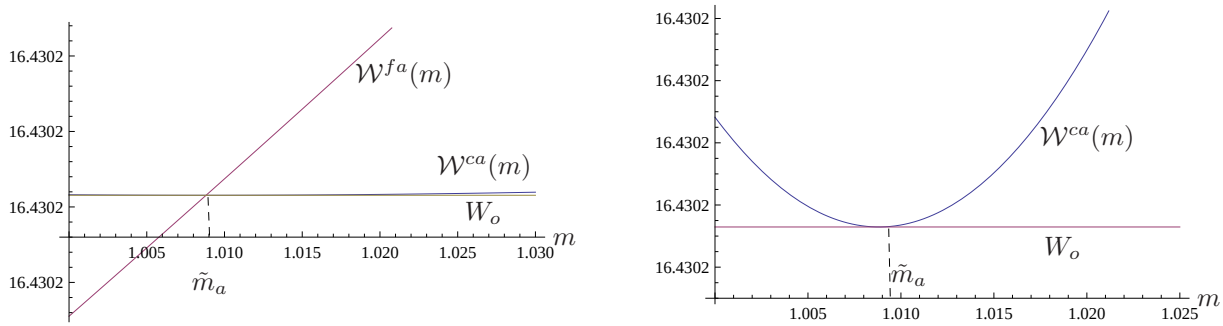


Figure 14: Welfare in Example 1 for autarky