

Stable Climate Coalitions (Nash) and
International Trade

Thomas Eichner
Rüdiger Pethig

CESIFO WORKING PAPER NO. 3915
CATEGORY 10: ENERGY AND CLIMATE ECONOMICS
AUGUST 2012

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the RePEc website:* www.RePEc.org
- *from the CESifo website:* www.CESifo-group.org/wp

Stable Climate Coalitions (Nash) and International Trade

Abstract

The basic model of the literature on self-enforcing international environmental agreements is a model of autarkic countries. We extend that model by international trade and investigate its impact on the performance of 'Nash' coalitions and on their stability, in particular, in a general equilibrium framework. First we characterize the performance of coalitions and non-coalition countries with regard to emissions and welfare and compare business as usual with the coalition-fringe scenario. In qualitative terms, the results in our free-trade model turn out to be the same as in the basic model for quadratic functional forms. In our model with international trade countries influence the terms of trade with their choice of policy and they make strategic use of that terms-of-trade effect. We find, however, that in the quadratic version of our model - as in the basic model - stable coalitions consist of no more than two countries. Finally, we explore the outcome of trade liberalization by moving from autarky to free trade. Although the coalition steps up its mitigation effort, world emissions rise which may be referred to as a 'green paradox of trade liberalization'. Trade liberalization turns out to be bad for the environment as well as for the coalition countries' welfare and the aggregate welfare of all countries; it reduces the range of profitable coalitions, and it even tends to hamper the formation of stable coalitions.

JEL-Code: C720, F500, Q500, Q580.

Keywords: sub-global climate coalition, international trade, trade liberalization, self-enforcing IEA.

Thomas Eichner
Department of Economics
University of Hagen
Universitätsstraße 41
Germany – 58097 Hagen
thomas.eichner@fernuni-hagen.de

Rüdiger Pethig
Department of Economics
University of Siegen
Hölderlinstraße 3
Germany – 57068 Siegen
pethig@vwl.uni-siegen.de

1 The problem

The massive reduction of global carbon emissions necessary to stabilize the world climate at safe levels cannot be achieved without an effective international environmental agreement (IEA). The first legally binding international climate agreement, the Kyoto Protocol, contains rather unambitious commitments for a small number of countries and therefore accomplishes only little more than global non-cooperation (Buchner et al. 2002). It phases out in 2012, and the prospects are bleak for reaching a new IEA with many signatories *and* substantial emission reduction commitments. The tedious practical negotiations on one hand and the serious global-change challenge on the other hand call for continued efforts to improve our understanding of the conditions for successful and effective IEAs.

The economic literature on IEAs since the early 1990s is based on the proposition that sovereign countries are reluctant to join an IEA unless it is in their self-interest. Therefore the concept of *self-enforcing* agreements is crucial which requires that no signatory has an incentive to leave the IEA and no non-signatory has an incentive to join it. The *basic model* of an IEA employed by Carraro and Siniscalco (1991), Hoel (1992), Barrett (1994)¹ and by many others since then is a static model of identical countries without international trade. Each country's domestic emissions generate domestic welfare that is decreasing at the margin and all countries' emissions create a welfare loss that is uniform across countries and increasing at the margin. The climate damage hits all countries and thus makes them interdependent but they are not interdependent via world markets and trade. One strand of the literature models climate coalitions² as Stackelberg leaders (e.g. Barrett 1994, Diamantoudi and Sartzetakis 2006, Rubio and Ulph 2006) and the other portrays them as Nash players along with all non-signatories (e.g. Carraro and Siniscalco 1991, Hoel 1992, Finus 2001). Most contributions are quite pessimistic about the stability of large IEAs. Nash modeling approaches find that stable coalitions consist of at most three countries, and the number of members of stable Stackelberg coalitions is not larger than four when negative emissions are ruled out.³ Hence in the *basic model* strong free-rider incentives prevent large

¹Barrett (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.

²In the present paper the terms IEA and (climate) coalition are synonymous because our exclusive focus is on a single coalition. Also, we take as equivalent the terms 'self-enforcing IEA' and 'stable (climate) coalition'.

³Barrett (1994) chooses the Stackelberg-coalition approach, and his simulations 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

stable coalitions and hence large gains from cooperation under both the Stackelberg and Nash assumption.

Motivated by relaxing the restrictive assumptions of the *basic model* and/or in search of conditions for more successful international cooperation, the basic model has been modified and extended in various directions (Finus 2003).⁴ However, we are not aware of analytical studies on the formation of IEAs that model in more detail the economies of individual countries and their economic interdependencies via trade.⁵ It is remarkable how little attention the coalition-formation literature paid so far to international trade although trade is at the core of the literature on strategic environmental policy with transboundary pollution. That topic was taken up in the seminal paper by Markusen (1975) who analyzes the optimal environmental and trade policy of one country when the other country is inactive. Maintaining the framework of transboundary pollution and perfect competition, Ludema and Wooton (1994) point out that environmental policy is used as a substitute for strategic trade policy, if the latter is not available and countries play Nash. Moreover, Rauscher (1997) and Lapan and Sikdar (2011) show in different settings that non-cooperative environmental policy of large countries is inefficiently lax under free trade and results in a 'race to the bottom'. Another (small) strand of literature (Copeland and Taylor 1995, Rauscher 2001, Hatzipanayotou et al. 2005 and Lapan and Sikdar 2011) examines whether or not trade is bad for the environment. Copeland and Taylor (1995) find in a Heckscher-Ohlin model with countries setting pollution quotas in a non-cooperative way that trade liberalization does not increase world pollution if countries are identical. However, none of these papers addresses the issue of coalition formation.

The main objective of the present paper is to extend the *basic model* of coalition formation which is essentially a model of the world economy with autarkic countries by investigating the impact of international trade on the performance of Nash climate coalitions⁶ and on their stability, in particular. To that end we add structure to the national economies by introducing a consumer good and fossil fuel that are produced in all countries, consumed by their representative consumers and traded on world markets. In autarky, all countries

infeasible to abate more emissions than are generated.

⁴It is not possible to provide an encompassing and balanced survey on extensions of the basic model of coalition formation. Major lines of analysis with sample references are: transfer schemes (Hoel and Schneider 1997, Fuentes-Albero and Rubio 2010), asymmetric countries (Barrett 2001, McGinty 2007, Pavlova and de Zeeuw 2012), uncertainty (Kolstad 2007), and altruism (van der Pol et al. 2012).

⁵Despite 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 may help support stable IEAs.

⁶In a companion paper we focus on Stackelberg climate coalitions and international trade. See Eichner and Pethig (2012).

are interdependent only through climate damage while in the free-trade regime the world markets for commodities and fuel create an additional interdependence which the coalition and all other countries may want to exploit by manipulating the terms of trade in their favor.

We first investigate how climate coalitions of given size perform in the free-trade regime. The focus is on comparing the performance of coalition and fringe countries as well as on comparing the coalition-fringe scenario with the business-as-usual scenario of global non-cooperation. For general non-parametric functions we prove (i) that coalition countries set tighter emission caps than fringe countries; that (ii) the welfare of coalition countries is lower than the welfare of fringe countries; and that (iii) total emissions are decreasing in the transition from business as usual to the coalition-fringe scenario. The results (i) - (iii) have also been established in the basic model for linear-quadratic functional forms⁷ by Carraro and Siniscalco (1991, 1993), Finus and Rundshagen (2001), De Cara and Rotillon (2001) and Rubio and Casino (2001).

Next, we turn to the issue of coalition stability. As noted above, in our free-trade model the fringe countries as well as the coalition have incentives to manipulate the terms of trade in their favor which is an option not available to autarkic countries. The question is whether that strategic element in the countries' choice of policy causes significant changes in the conditions for the formation of stable coalitions. We find that for the quadratic version of our model the size of stable coalitions is at most two which is again the same result as in the basic model. Thus we prove that the direct effects of the coalition's and the fringe countries' emission caps on climate damage turn out to be stronger than the indirect terms-of-trade effects, i.e. the effects of their emission caps on the fuel price expressed in terms of the consumption good.

These findings suggest that international trade does not make a difference in qualitative terms. Yet in order to fully understand the role of international trade in the context of coalition formation, we also explore the outcome of trade liberalization by comparing the regimes of autarky and free trade. We show that in the transition from autarky to free trade the coalition sets a tighter cap and the fringe countries set laxer caps. Although the coalition steps up its mitigation effort, world emissions rise which may be referred to as a 'green paradox of trade liberalization'. We also show that opening the borders is welfare reducing for coalition countries but may either enhance or reduce the welfare of fringe countries depending on the size of the coalition. Trade liberalization is not only bad for the environment but it also diminishes the sum of all countries' welfare, it reduces the range of profitable coalitions, and it even tends to hamper the formation of stable coalitions

⁷Rubio and Ulph (2005) call that parametric model the workhorse for analyzing stable coalitions.

in the following way. If in the autarky regime of some given economy a stable two-country coalition exists (which is the only size of a stable coalition), then under certain conditions that coalition is not stable anymore in the free-trade regime; and if there is no stable coalition in autarky, there will be no stable coalition in the free-trade regime either.

The paper is organized as follows: Section 2 introduces and describes the formal model. Section 3 studies the business-as-usual scenario which is the same in the regimes of free trade and autarky due to the assumption of symmetric countries. Section 4 focuses on the coalition-fringe scenario in the free-trade regime. First, it analyzes the impact of given climate coalitions on emissions and welfare and then it deals with the existence and size of self-enforcing IEAs. Section 5 clarifies the role of international trade by investigating the impact of trade liberalization on emissions, welfare and the self-enforcing IEAs. Section 6 concludes.

2 The model

The world economy consists of n identical countries. Each country produces two consumer goods. The first is a standard composite consumption good, called *good X* (quantity x_i) and the second is a fossil energy carrier (quantity e_i), e.g. gas or coal extracted from domestic fossil reserves. We refer to that good simply as *fuel*. Each country's production technology is represented by the production possibility frontier⁸

$$x_i^s = T(e_i^s) \quad i = 1, \dots, n, \quad (1)$$

where the function T is decreasing and 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.

$$V(e_i^d) + x_i^d - D\left(\sum_j e_j^d\right) \quad i = 1, \dots, n \quad (2)$$

is the utility of the representative consumer of country i . The function V is increasing and strictly concave and the climate damage function D is increasing and strictly convex in its arguments. Fuel units are chosen such that e_i^d denotes both fuel demanded by consumer i and carbon emissions from burning fuel. The supply constraints

$$\sum_j x_j^s = \sum_j x_j^d \quad \text{and} \quad \sum_j e_j^s = \sum_j e_j^d \quad (3)$$

⁸The superscripts s and d indicate quantities supplied and demanded, respectively. Functions are upper case letters and subscripts attached to those letters denote partial derivatives. The notation $\sum_j z_j$ is short for $\sum_{j=1}^{j=n} z_j$.

allow for international trade in both commodities.

There are world markets for good X (price $p_x \equiv 1$) and for fuel (producer price p_e). All governments regulate emissions by putting a binding cap e_i on domestic fuel consumption (= emissions):

$$e_i^d = e_i \quad i = 1, \dots, n. \quad (4)$$

Government i implements its cap by auctioning the amount e_i of emission permits with π_i being the permit price in country i . Thus the consumer price of fuel is $p_e + \pi_i$. With that institutional setup of demand-side regulation, the policy variable of government i is the cap e_i . In the following sections we will focus on welfare maximizing governments. But for expository purposes we first take these caps to be arbitrarily fixed.

There is one producer in each country i who maximizes profits $\Pi^i := x_i^s + p_e e_i^s$ subject to (1). The straightforward implication of the first-order condition is

$$p_e = -T'(e_i^s) \quad i = 1, \dots, n, \quad (5)$$

which readily implies $e_i^s = e_j^s$ for all $i, j = 1, \dots, n$. Since all countries are alike and all producers face the same (relative) price p_e , the output bundle (e_i^s, x_i^s) is the same in all countries.

Consumer i ignores her contribution to climate damage and maximizes over e_i^d her 'non-climate utility' $V(e_i^d) + x_i^d$ subject to her budget constraint $x_i^d + (p_e + \pi_i)e_i^d = y_i$, where

$$y_i := p_e e_i^s + T(e_i^s) + \pi_i e_i^d \quad (6)$$

is country i 's income. From the corresponding first-order condition follows immediately

$$p_e + \pi_i = V'(e_i^d). \quad (7)$$

In the economy (1) - (7) there exists a unique competitive equilibrium for every set of binding emission caps (e_1, \dots, e_n) . To see that, combine (3), (4) and (5) to obtain

$$e_i^s = e_j^s = \frac{\sum_k e_k}{n} \quad i, j = 1, \dots, n. \quad (8)$$

Thus the equilibrium prices p_e and π_i are determined by the caps (e_1, \dots, e_n) via (4), (5), (7) and (8), and the equilibrium consumption of good X turns out to be

$$x_i^d = T\left(\frac{\sum_j e_j}{n}\right) - \left(\frac{\sum_j e_j}{n} - e_i\right) T'\left(\frac{\sum_j e_j}{n}\right). \quad (9)$$

For the benefit of more specific results, we will specify later in some parts of the paper the functions T , V and D from (1) and (2) by the following quadratic functional forms⁹

$$T(e_i^s) = \bar{x} - \frac{\alpha}{2} (e_i^s)^2, \quad V(e_i^d) = \beta e_i^d - \frac{\gamma}{2} (e_i^d)^2, \quad D\left(\sum_j e_j\right) = \frac{\delta}{2} \left(\sum_j e_j^d\right)^2, \quad (10)$$

where \bar{x} , α , β , γ and δ are positive constants.

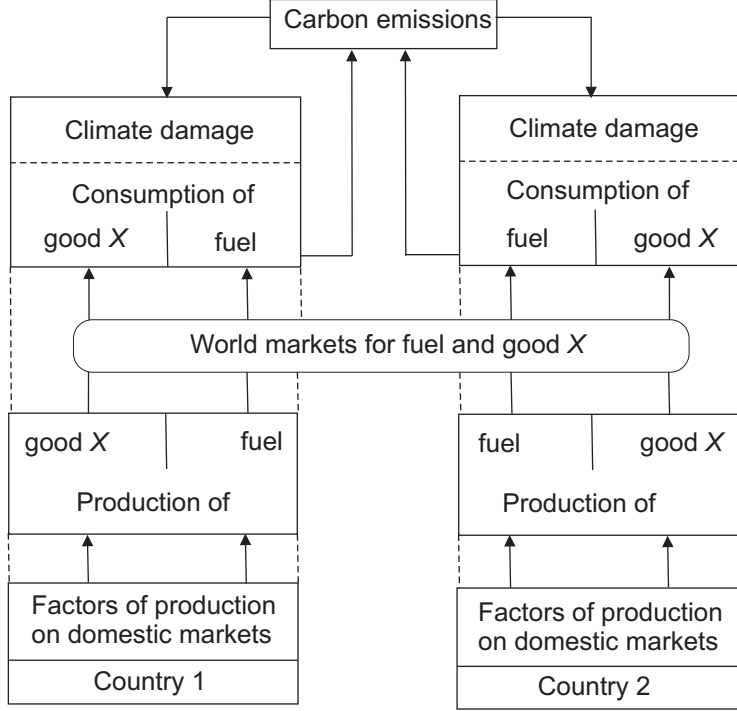


Figure 1: The structure of the model

Figure 1 illustrates and summarizes the structure of the model and points, in particular, to the prominent role of international trade in fuel and consumption goods. As Figure 1 shows, 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. In practice climate regulation is applied to the consumers' energy demand as well as to the demand of energy-consuming industries, for the latter of which the EU emission trading scheme is a prominent example. To keep the analysis tractable we have refrained from modeling fuel as an intermediary input in the production of good X . Yet our simplification captures the thrust of emission regulation, because more stringent emission caps require raising the domestic price of fuel consumption whether fuel consumption is

⁹Let \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 $r_e + r_x = \bar{r}$. α_e, α_x are positive constants. The quadratic transformation function in (10) is straightforward from these three equations when setting $\bar{x} := \alpha_x \bar{r}$ and $\alpha := \alpha_x \alpha_e$. The production functions imply that α measures the progressivity of the extraction costs.

regulated on the level of industries or consumers. Since all allocative changes work through variations in relative prices, they are qualitatively similar for both types of regulation.

From (2), (5) and (9) follows the equilibrium welfare of country i ,

$$W^i(e_1, \dots, e_n) = V(e_i) - G^i(e_1, \dots, e_n), \quad (11)$$

$$\text{where } G^i(e_1, \dots, e_n) := \underbrace{D\left(\sum_j e_j\right)}_{\substack{\text{climate change} \\ \text{interdependency}}} + \underbrace{\left(\frac{\sum_j e_j}{n} - e_i\right) T'\left(\frac{\sum_j e_j}{n}\right) - T\left(\frac{\sum_j e_j}{n}\right)}_{\text{world market interdependency}}.$$

Equation (11) highlights the dependence of country i 's welfare on all countries' emission cap policies through two different channels, the climate damage channel and the international trade channel. Throughout the paper we will restrict our attention to that class of welfare functions W^i which are strictly concave in e_i and attain an interior maximum with respect to e_i in the relevant domain of $(n-1)$ -tuples $(e_1, \dots, e_{i-1}, e_{i+1}, \dots, e_n)$. We show in the Appendix A, that these properties of W^i can be secured under mild (sufficient) conditions.

3 Business as usual (BAU) as a benchmark

In this section all governments are assumed to 'behave Nash'. That is, given e_j for all $j \neq i$, government i chooses that cap e_i which maximizes its country's welfare $W^i(e_1, \dots, e_n)$ from (11) for predetermined caps $(e_1, \dots, e_{i-1}, e_{i+1}, \dots, e_n)$. Country i 's best-reply function is implicitly given by the first-order condition

$$W_{e_i}^i = V'(e_i) - \left(\frac{\sum_j e_j}{n} - e_i\right) \frac{T''\left(\frac{\sum_j e_j}{n}\right)}{n} + T'\left(\frac{\sum_j e_j}{n}\right) - D\left(\sum_j e_j\right) = 0 \quad (12)$$

and explicitly by

$$e_i = \tilde{R}(\sigma_i) \quad \text{with} \quad \tilde{R}_{\sigma_i} := \frac{de_i}{d\sigma_i} = -\frac{nG_{e_i e_i}^i + T''}{nG_{e_i e_i}^i - nV''} \in]-1, 0[\quad \text{and} \quad \sigma_i := \sum_{j \neq i} e_j. \quad (13)$$

Since all countries are alike, the function \tilde{R} is the same for all countries, and its property $\tilde{R}_{\sigma_i} \in]-1, 0[$ characterizes all countries' emission caps as strategic substitutes. Under these conditions the Nash equilibrium (e_{1o}, \dots, e_{no}) is unique and satisfies $e_{1o} = e_{2o} = \dots = e_{no} =: e_o$. Recall from (5) that the fuel supplies are the same in all countries. Therefore it is obvious from (3), (4) and the uniformity of equilibrium caps that no international trade takes place in spite of open borders. Due to the symmetry assumption, the first-order condition turns into

$$V'(e_o) = -T'(e_o) + D'(ne_o) \quad (14)$$

which shows that the BAU emissions cap e_o is such that a country's marginal benefit of consumption, $V'(e_o)$, equals its marginal rate of transformation, $-T'(e_o)$, and its marginal climate damage, $D'(ne_o)$.

While the above characterization of the non-cooperative n -country game is straightforward, we now aim to convert the n -country game into a non-cooperative game with two groups of countries. The purpose of this conversion is to develop analytical tools as well as a benchmark for our later analysis of (stable) climate coalitions under the Nash assumption. Specifically, we lump together the first m countries, $2 \leq m < n$, in one group, denoted group $C = \{1, 2, \dots, m\}$, and collect the remaining countries in another group, denoted group $F = \{m + 1, \dots, n\}$. We introduce that grouping here to prepare for the later study of a scenario in which all countries in group C commit to a joint cooperative climate policy (with C for coalition) whereas all countries in group F (with F for fringe) continue to act non-cooperatively. It is important to emphasize, however, that right now the grouping is purely formal, that is, each member of either group still acts according to its best-reply function (13).

As we know that all countries in one and the same group choose the same cap in the Nash equilibrium, we can simplify the formal analysis w.l.o.g. by setting $e_i = e_c$ for all $i \in C$ and $e_i = e_f$ for all $i \in F$. Correspondingly, the best-reply functions (13) are turned into $e_h = \tilde{R}^h(\sigma_h)$ for $h = c, f$ where $\sigma_c := (m - 1)e_c + (n - m)e_f$ and $\sigma_f := me_c + (n - m - 1)e_f$. We show in the Appendix B, that there are best-reply functions $R^h, h = c, f$, equivalent to the best-reply functions \tilde{R}^h , satisfying¹⁰

$$me_c = R^c[(n - m)e_f, m] \quad \text{and} \quad (n - m)e_f = R^f(me_c, m), \quad (15)$$

$R_{me_c}^f \in] - 1, 0[$ and $R_{(n-m)e_f}^c \in] - 1, 0[$. By construction of the functions R^c and R^f it is true that $e_c = e_f = e_o$ implies $me_o = R^c[(n - m)e_o, m]$ and $(n - m)e_o = R^f(me_o, m)$ for all $m \in [0, n]$. That is, in the two-group Nash equilibrium $[me_o, (n - m)e_o]$ all countries' emission caps are the same as in the game with n non-cooperative countries, and that holds for any size m of group C .

To sum up, we have turned the non-cooperative game of n identical countries with strategies e_i ($i = 1, \dots, n$) into a non-cooperative game of two players C and F whose strategies are me_c and $(n - m)e_f$, respectively, without changing the 'original' n -country game in its substance.

Figure 2 illustrates the two-group Nash equilibrium as the point of intersection of the

¹⁰It is analytically convenient to treat m as a real number. However, when we later will investigate the issue of coalition stability, we will take into consideration that m is a positive integer.

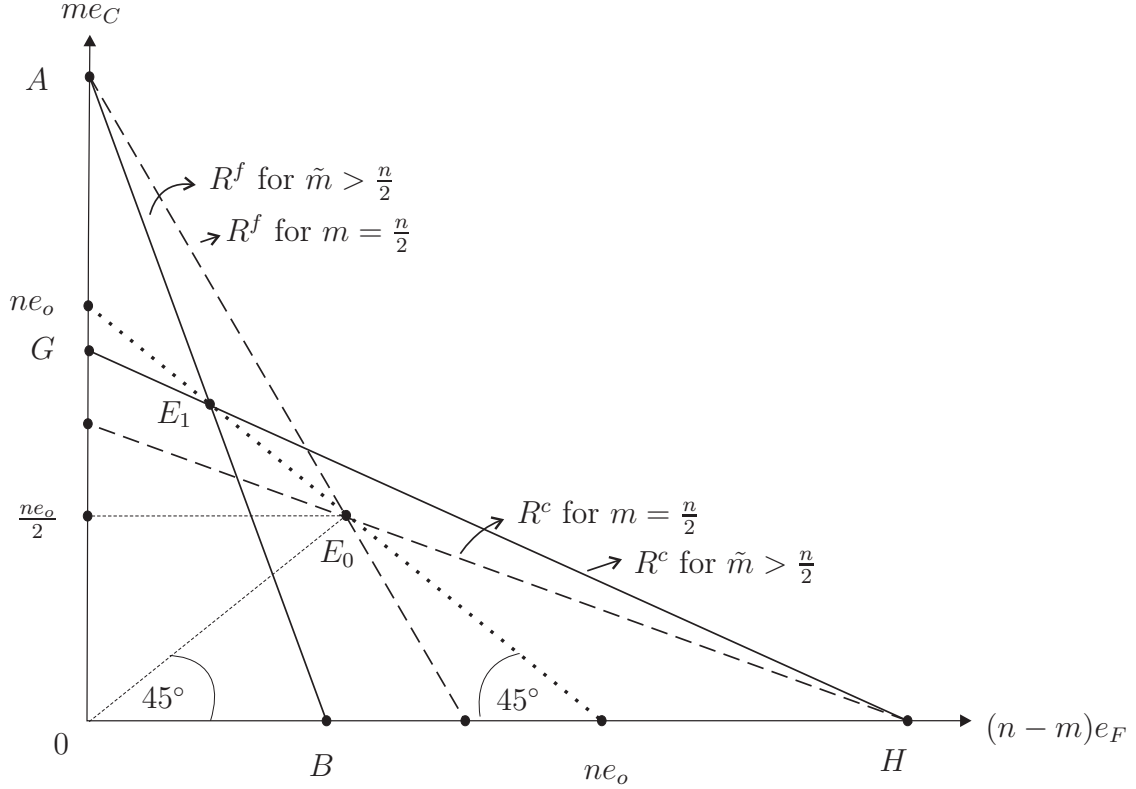


Figure 2: The n -country game as a game of two groups of countries

graphs¹¹ of the best-reply functions (15) for two alternative values of m . The dashed lines are the best-reply curves for $m = n/2$. Their intersection point is E_0 , and the corresponding group equilibrium emissions are $me_c = (n - m)e_f = (n/2)e_o$. The solid straight lines in Figure 2 represent the best-reply curves for some value $\tilde{m} > n/2$. Their intersection point is E_1 , and the corresponding equilibrium emissions $\tilde{m}e_o$ and $(n - \tilde{m})e_o$ obviously satisfy $\tilde{m}e_o > (n/2)e_o > (n - \tilde{m})e_o$ because $\tilde{m} > n/2$ by presupposition. Note also that total emissions are ne_o in both equilibrium points, E_0 and E_1 . More generally, the dotted straight 45° line (with points ne_o on both axes) is the locus of all Nash equilibria. With increasing m the equilibrium point on the dotted line moves closer toward the vertical axis. We summarize our results in

Proposition 1. *Suppose all n countries pursue non-cooperative climate policies.*

- (i) *There exists a unique Nash equilibrium $(\underbrace{e_o, \dots, e_o}_{n\text{-times}})$ (= business as usual). The uniform emission cap e_o of all countries is binding and international trade does not take place*

¹¹We have drawn the graphs of the best-reply functions (15) as straight lines for convenience of exposition. They are straight lines, in fact, in the parametric version of the model introduced in equation (10). As shown in Figure 2, if one employs the parametric functions (10) the best-reply curve of group C [group F] for $\tilde{m} > n/2$ deviates from the curve of that group for $m = n/2$ via rotation around its point on the horizontal [vertical] axis away from [toward] the origin.

(although the borders are open).

- (ii) Equivalent to the n -country game is a game of two groups of countries, say the groups $C = \{1, 2, \dots, m\}$ and $F = \{m + 1, \dots, n\}$ with $m \in \{1, \dots, n\}$, whose strategies are the emissions me_c and $(n - m)e_f$, respectively. The corresponding Nash equilibrium $[me_{co}, (n - m)e_{fo}]$ exists, is unique and satisfies $e_{co} = e_{fo} = e_o$ for all $m \in \{1, \dots, n\}$ (and hence is identical to the equilibrium in Proposition 1(i)).

Our presentation of the no-cooperation scenario as a game of two groups all of whose members act non-cooperatively is not particularly interesting in its own right. However, it will turn out to be a useful tool for the subsequent analysis of sub-global cooperation.

4 The coalition-fringe scenario

In this section we assume that the number of countries, m , in group C is exogenous and that all m countries have committed to cooperate in climate policy, while all other countries $i = m + 1, \dots, n$, called fringe countries, continue to abstain from cooperation. Each fringe country $i \in F$ plays Nash against all other fringe countries and against group C , but the climate coalition C now acts as a single player whose payoff is the coalition members' aggregate welfare and who plays Nash against all fringe countries. We will first characterize - and compare with BAU - the emissions of all countries in the coalition-fringe equilibrium (Section 4.1) and then the welfare levels (Section 4.2); we also investigate the welfare implications of changes in the coalition size by means of numerical examples.

4.1 Impact of the climate coalition on emissions

We continue to use the notation¹² $e_i = e_c$ for all $i \in C$ and $e_i = e_f$ for all $i \in F$ and define $e := me_c + (n - m)e_f$ for further notational relief. Based on (11) we then express the coalition's aggregate welfare as

$$\sum_{j \in C} W^j[me_c, (n - m)e_f, m] = mV\left(\frac{me_c}{m}\right) - mG^c[me_c, (n - m)e_f, m], \quad (16)$$

¹²Setting $e_i = e_c$ for all $i \in C$ is admissible here, but for a different reason as in Section 3. Group C now maximizes the aggregate welfare of its member countries. Since all countries are identical, equal treatment is a necessary condition for maximizing aggregate welfare.

where $G^c[\cdot] = D(e) - T\left(\frac{e}{n}\right) + \left(\frac{e}{n} - \frac{me_c}{m}\right)T'\left(\frac{e}{n}\right)$. As shown in Appendix C, maximizing (16) with respect to me_c yields the coalition's best-reply function¹³

$$me_c = \hat{R}^c[(n-m)e_f, m] \quad \text{with} \quad \hat{R}_{(n-m)e_f}^c \in]-1, 0[. \quad (17)$$

Since we keep portraying fringe countries as non-cooperative Nash players, their 'aggregate' action is still described by the best-reply function $(n-m)e_f = R^f(me_c, m)$ from (15). Consequently, a coalition-fringe Nash equilibrium is a tuple $[me_c^*, (n-m)e_f^*]$ satisfying $me_c^* = \hat{R}^c[(n-m)e_f^*, m]$ from (17) as well as $(n-m)e_f^* = R^f(me_c^*, m)$ from (15). Owing to $\hat{R}_{(n-m)e_f}^c \in]-1, 0[$ and $R_{me_c}^f \in]-1, 0[$, the Nash equilibrium is unique.¹⁴

In order to determine the sign of the difference $e_f^* - e_c^*$ we write the first-order condition of a fringe country after some rearrangement of terms as

$$V'(e_f) - \frac{m}{n^2}(e_c - e_f)T''\left(\frac{e}{n}\right) = -T'\left(\frac{e}{n}\right) + D'(e). \quad (18)$$

Analogously, from maximizing (16) with respect to me_c we obtain the first-order condition

$$V'(e_c) - \frac{m(n-m)}{n^2}(e_f - e_c)T''\left(\frac{e}{n}\right) = -T'\left(\frac{e}{n}\right) + mD'(e). \quad (19)$$

An interior Nash equilibrium $[me_c^*, (n-m)e_f^*]$ requires (18) and (19) to hold simultaneously. Subtracting (18) from (19) yields

$$V'(e_c^*) - V'(e_f^*) - \frac{m(n-m+1)}{n^2}(e_f^* - e_c^*)T''\left(\frac{e^*}{n}\right) = (m-1)D'(e^*). \quad (20)$$

Satisfying (20) requires its left-hand side to be positive. Owing to $V' > 0$, $V'' < 0$ and $T'' < 0$ this is secured, if and only if $e_c^* < e_f^*$. As all countries supply the same amount of fuel, an immediate implication of $e_c^* < e_f^*$ is that in Nash equilibrium the coalition exports fuel and all fringe countries import fuel.

Next, we compare the emissions e_c^* and e_f^* with the BAU emissions (per country) e_o . The key for the comparison is the observation that the tuples $[me_c^*, (n-m)e_f^*]$ and $[me_o, (n-m)e_o]$ both satisfy the best-reply function R^f . To put it graphically, both are points on the graph of R^f , say on the line AB in Figure 2, where the point E_1 represents the BAU equilibrium $[me_o, (n-m)e_o]$. $e_c^* < e_f^*$ implies that $[me_c^*, (n-m)e_f^*]$ is a point on the segment E_1B . That observation immediately yields $e_c^* < e_o < e_f^*$. To determine the sign of $ne_o - e^*$ observe that $e = me_c + (n-m)e_f = me_c + R^f(me_c)$ and $(de/dme_c) = 1 + R_{me_c}^f$. From $R_{me_c}^f \in]-1, 0[$ follows $1 + R_{me_c}^f > 0$ which means, in turn, that if me_c is reduced from me_o to me_c^* , total emissions e decline from ne_o to $e^* < ne_o$. We summarize these findings in

¹³ \hat{R}^c in (17) differs from R^c in (15) because the former relates to the cooperative group C and the latter to the non-cooperative group C .

¹⁴In graphical terms, there is one and only one intersection point of the graphs of the best-reply functions \hat{R}^c and R^f (which we assume to be an interior solution).

Proposition 2. *In the coalition-fringe Nash equilibrium $(me_c^*, \underbrace{e_f^*, \dots, e_f^*}_{(n-m)\text{-times}})$ the emissions satisfy $e_c^* < e_f^*$. Compared with the BAU scenario we find $e_c^* < e_o < e_f^*$ and $me_c^* + (n - m)e_f^* < ne_o$.*

The core message of Proposition 2 is that the climate coalition is successful in the sense that all its member countries step up their mitigation effort ($e_c^* < e_o$) and that the coalition succeeds in bringing down total world emissions ($e^* < ne_o$). The inequalities $e_c^* < e_o < e_f^*$ and $e^* < ne_o$ mean that the fringe countries respond to the coalition's climate policy by lowering their own mitigation efforts thus thwarting - not fully but to some extent - the coalition's increased effort of fighting climate change. Carbon does leak from the coalition to the fringe but the 'free riding' fringe does not fully compensate or overcompensate the coalition's effort of curbing world emissions. The coalition's emission reduction is greater than the expansion of emissions its policy causes in the fringe countries. In the language of carbon leakage we find that subglobal climate coalitions induce carbon leakage at a rate that is positive but less than 100%.

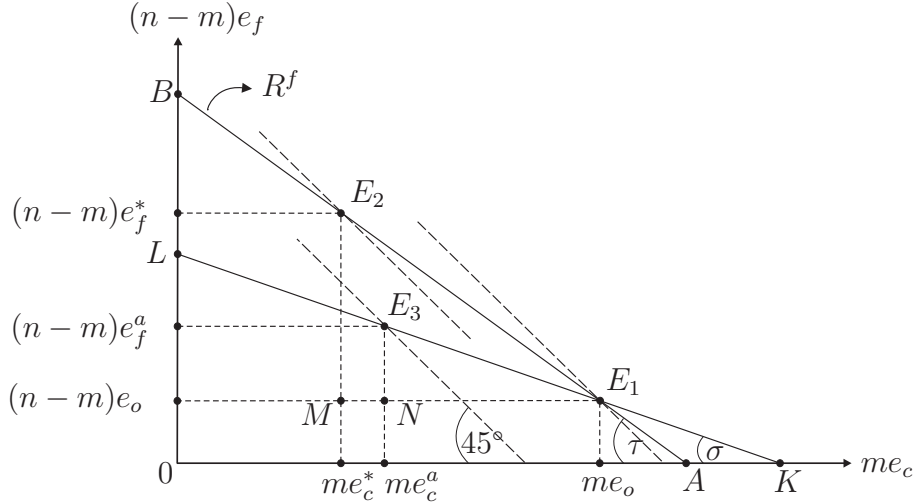


Figure 3: Emissions in BAU (E_1) and in the coalition-fringe equilibrium with (E_2) and without trade (E_3)

Proposition 2 is illustrated in Figure 3. The curve AB is the same as in Figure 2, namely the graph of the fringe group's best-reply function R^f . We ignore the line KL in Figure 3 right now because it relates to an analysis to be conducted later in Section 5. The points E_1 and E_2 on the curve AB are supposed to represent the BAU equilibrium and the coalition-fringe equilibrium, respectively. To avoid overloading Figure 3 we have suppressed the graphs of the best-reply functions \hat{R}^c and R^c . What is important about them is their point of intersection with the graph of R^f . In line with Proposition 2 we have positioned the equilibrium points E_1 and E_2 such that $e_c^* < e_o < e_f^*$. Note also that Figure 3 displays three

dotted straight lines passing through E_1 , E_2 and E_3 , respectively. By construction, they have a negative slope of 45° such that world emissions $e = me_c + (n - m)e_f$ are constant on each of them and the level of world emissions increases with the distance of the line from the origin. Thus the dotted lines passing through E_1 and E_2 illustrate the inequality $e^* < ne_o$. Figure 3 also illustrates leakage rates. The leakage rate of sub-global cooperation (transition from BAU to the coalition-fringe scenario) is equal to $(E_2M/ME_1) < 1$.

So far we have studied the emissions in the coalition-fringe scenario (compared to BAU) for some given size $m \in \{2, \dots, n - 1\}$ of the coalition. It is clear (but has been suppressed so far) that the coalition fringe equilibrium caps and the countries' corresponding welfare depend on the size of the coalition. For the sake of the formal analysis we write $e_c^* = \mathcal{E}^c(m)$, $e_f^* = \mathcal{E}^f(m)$ and

$$\begin{aligned} \mathcal{W}^c(m) &:= W^j [m\mathcal{E}^c(m), (n - m)\mathcal{E}^f(m)] \quad \text{for } j \in C, \\ \mathcal{W}^f(m) &:= W^j [m\mathcal{E}^c(m), (n - m)\mathcal{E}^f(m)] \quad \text{for } j \in F. \end{aligned}$$

We know already that for $m = 1$ the caps are $\mathcal{E}^c(1) = \mathcal{E}^f(m) = e_o$ and that for $m = n$ the caps $\mathcal{E}^c(n)$ are at their socially optimal level which is smaller than e_o .¹⁵ If we start from $m = 2$ and successively increase the coalition size m , we know that the emissions per country converge toward the socially optimal emissions. More specific results are proved in the Appendix C and summarized in

Proposition 3. *Consider coalition fringe equilibria $[m\mathcal{E}^c(m), (n - m)\mathcal{E}^f(m)]$ for alternative coalition sizes m , $1 \leq m < n$. Total emissions $m\mathcal{E}^c(m) + (n - m)\mathcal{E}^f(m)$ are strictly decreasing in m and the fringe countries' emissions $\mathcal{E}^f(m)$ are strictly increasing in m .*

The result that total emissions decline monotonely with increasing coalition size from their BAU level to the socially optimal level conforms to one's intuition. Climate damage can be mitigated the more, the larger is the coalition. It is rather unexpected, however, that $\mathcal{E}^f(m)$ grows when the number of fringe countries declines. The smaller is the group of fringe countries the more profitable free riding is for the individual fringe country and the greater is carbon leakage per fringe country.¹⁶

¹⁵The socially optimal level of emissions is the level emissions would have in the grand coalition where all climate damage externalities are fully internalized.

¹⁶Although the sign of the derivative \mathcal{E}_m^c is ambiguous, some further information beyond Proposition 3 can be gained by observing that $\mathcal{E}^c(1) = e_o$ and $\mathcal{E}^c(n) < e_o$. Therefore, $\mathcal{E}^c(m)$ tends to be decreasing in m although that mapping need not be monotone.

4.2 Impact of the climate coalition on welfare

The transition of the countries in group C from non-cooperation to cooperation has significant welfare implications for all countries, inside and outside the coalition. Unfortunately, the complexity of the model allows for limited analytical results only, but important insights can be gained through graphical and numerical analysis.

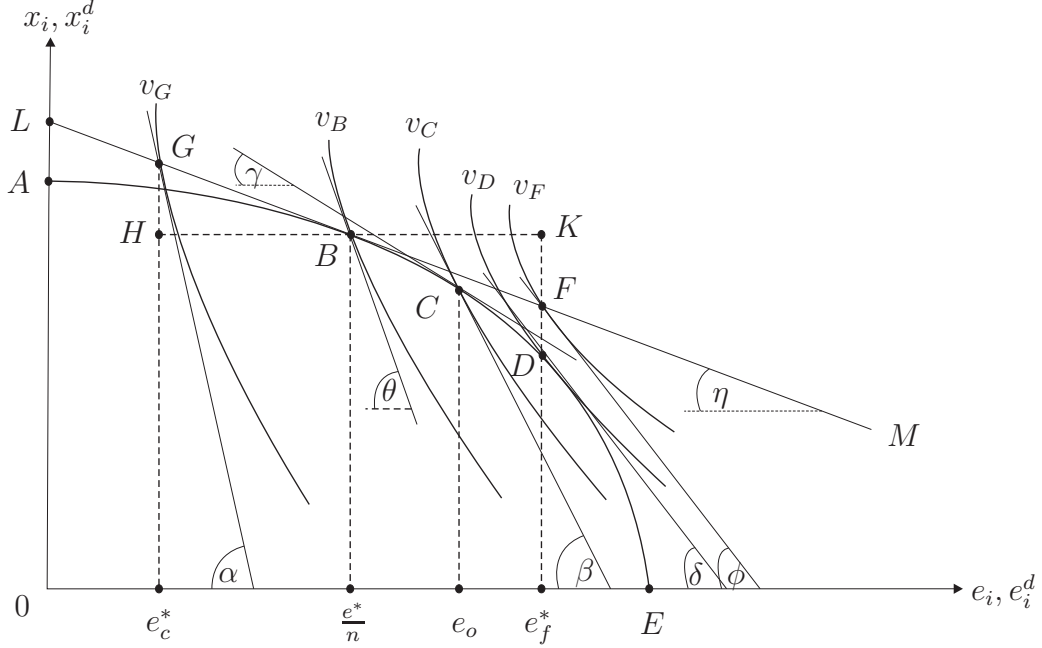


Figure 4: Comparisons of welfare net of climate damage

In Figure 4 the curve $ABCDE$ is the graph of the transformation function (1) of an individual country. The indifference curves v_G, v_B etc. indicate the individual countries' welfare associated with the consumption bundle (e_i^d, x_i^d) . As the welfare loss of climate damage enters the utility function (2) in an additive way, the indifference curves in Figure 4 are invariant with respect to climate damage. Since all countries are alike, the transformation curve and the indifference curves are the same for all countries.

Consider first the point D in Figure 4 where the marginal rate of transformation equals the marginal rate of substitution in consumption. D clearly is each country's production and consumption point in the no-policy scenario (no regulation, no trade, fuel price = $\tan \delta$). A point on the transformation curve to the left of D , such as Point C is each country's production and consumption point in the BAU Nash equilibrium. As in the no-policy case, there is no trade, but cap-and-trade schemes are installed, and the uniform permit price is equal to $\pi = \tan \beta - \tan \gamma > 0$. It is also straightforward from Figure 4 that the BAU emissions (total and per country) are smaller than the emissions in the no-policy regime.

Having depicted the BAU scenario as a benchmark we are ready to illustrate the

coalition-fringe Nash equilibrium in Figure 4. In that equilibrium each country produces the same amount of fuel, e^*/n , that is smaller than e_o - as established in Proposition 2. Hence the common production point is a point such as point B on the transformation curve. The uniform producer price of fuel is $\tan \eta$, and that price represents the terms of international trade. Consequently the consumption points of all countries must lie on the straight line LM that is tangent to the transformation curve at point B . Since $e_c^* < e_f^*$ holds for fuel consumption (Proposition 2), all countries in group C export fuel - and hence their consumption point is a point such as G northwest of point B - whereas all countries in group F import fuel - and hence their consumption point is a point such as F southeast of point B . For convenience of exposition, Figure 4 shows the special case where n is an even number ($n > 2$) and $m = n/2$. Then it follows, as drawn in Figure 4, that the fuel export HB [commodity import GH] of each country in group C is equal to the fuel import BK [commodity export KF]¹⁷ of each country in group F .¹⁸

In our discussion of Figure 4 we have made use of the indifference curves v_G, v_B etc. to identify consumption points as points of tangency between some indifference curve and an appropriately chosen budget line. A closer look at the welfare levels represented by those indifference curves allows assessing the welfare changes of individual countries in the transition from BAU to the coalition-fringe scenario. Although Figure 4 can only be used for comparing welfare levels net of climate damage (non-climate welfare), comparisons of non-climate welfares correctly indicate rankings of total welfares because climate welfares are the same for all countries. Therefore we conclude from inspecting Figure 4 that in all countries the non-climate welfare is higher in the no-policy scenario than in the BAU equilibrium ($v_D > v_C$). Also, the changes in non-climate welfare induced by the transition from BAU to the coalition-fringe scenario differ markedly between fringe and coalition countries because their consumption points differ. In the coalition-fringe equilibrium the coalition countries' non-climate welfare v_G is smaller than the fringe countries' non-climate welfare v_F such that for any given coalition size m fringe countries are better off than coalition countries. The conclusions are summarized in

Proposition 4. *The transition from the BAU equilibrium $[me_o, (n - m)e_o]$ to the coalition-fringe equilibrium $[me_c^*, (n - m)e_f^*]$ leads to the following welfare changes:*

- (i) *The climate welfare rises in all countries by the amount $D(ne_o) - D(e^*) > 0$.*
- (ii) *The non-climate welfare*

¹⁷We have plotted the point F in Figure 4 exactly above point D to avoid clutter. The general feature that needs to be observed is $e_o < e_f^*$.

¹⁸For $m \neq n/2$ it would be still true that $(e^*/n) < e_o < e_f^*$ as drawn in Figure 4. However, for $m > n/2$ [$m < n/2$] the triangle GHB would be larger [smaller] than the triangle FKB .

- declines in coalition countries (by the amount $v_C - v_G$ in Figure 4) and
- either rises in fringe countries (by the amount $v_F - v_C$ in Figure 4) or declines in fringe countries by less than in coalition countries.¹⁹

An immediate implication of Proposition 4 is that for any given coalition size, the total welfare of fringe countries is greater than the total welfare of coalition countries. Up to now we have characterized the countries' welfare changes induced by the transition from BAU to a coalition fringe regime with some fixed coalition size m and the difference between the welfare of coalition and fringe countries in the coalition fringe regime with some fixed coalition size m . We proceed by investigating how the countries' welfare responds to exogenous variations in the coalition size. In the Appendix D we prove

Proposition 5.

- (i) *The welfare of a fringe country, $\mathcal{W}^f(m)$, is strictly increasing in m .*
- (ii) *The welfare of a coalition country, $\mathcal{W}^c(m)$, is strictly decreasing in m for $m = 1$.*
- (iii) *The welfare of a coalition country in BAU is smaller than that of a coalition country in the grand coalition ($\mathcal{W}^c(n) > \mathcal{W}^c(1)$).*

Proposition 5(i) states that fringe countries are better off with increasing coalition size, that is they extend their free ride on the coalition's mitigation effort the more the larger is the coalition. In contrast, the welfare of coalition countries is not monotone increasing in the coalition size. According to Proposition 5(ii) the members of small coalitions may suffer a welfare loss compared to BAU because $\mathcal{W}_m^c(m) < 0$ for $m = 1$. But with increasing coalition size their welfare is above BAU because it equals the socially optimal level if $m = n$ (Proposition 5(iii)). Altogether the Propositions 5(ii) and 5(iii) suggest that after having taken a dip the welfare of coalition countries rises with increasing m .

In order to obtain more informative answers (under less general assumptions, though), we resort to simulations making use of the parametric functions (10). Our first example (Example 1) is made up of the parameter values $n = 100, \alpha = 1000, \beta = 10000, \gamma = 5000$ and $\delta = 1$ and is illustrated in Figure 5. The left panel of Figure 5 depicts the graphs of the welfare functions \mathcal{W}^c and \mathcal{W}^f . Both graphs clearly have a point in common at $m = 1$ (business as usual) and in the left panel of Figure 5 both seem to be positively sloped for all $m \geq 1$. That holds, indeed, for \mathcal{W}^f (Proposition 5(i) but not for \mathcal{W}^c (Proposition 5(ii)).

¹⁹Observe that Figure 4 is drawn for $m = \frac{n}{2}$. For small m the point F may lie northwest of point C .

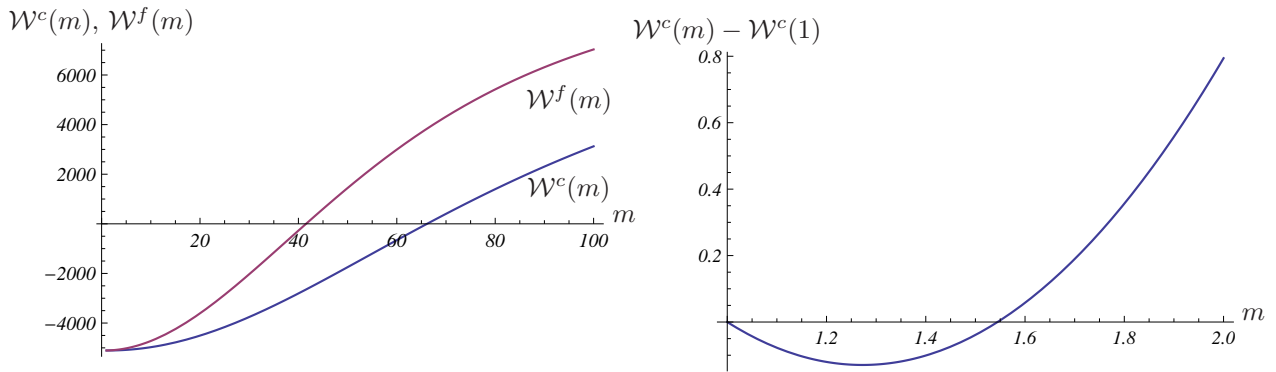


Figure 5: Welfare for coalition and fringe countries (Example 1)

A closer look at the graph of the function \mathcal{W}^c in the neighborhood of $m = 1$ in the right panel of Figure 5 proves that $\mathcal{W}_m^c(m) < 0$ for $m = 1$ is satisfied for Example 1, as required by Proposition 5(ii). The enlarged section of the graph of \mathcal{W}^c in the right panel of Figure 5 also shows that $\mathcal{W}^c(2) > \mathcal{W}^c(1)$. Hence in Example 1 the coalition countries' welfare exceed their BAU welfare for every coalition size $m \geq 2$ and we infer from Figure 5 that the welfare of all countries is increasing in the coalition size for all $m \geq 2$.²⁰ Moreover, for all $m > 1$ the fringe countries' welfare is greater than that of coalition countries - as already stated in Proposition 4(ii) in a more general context.

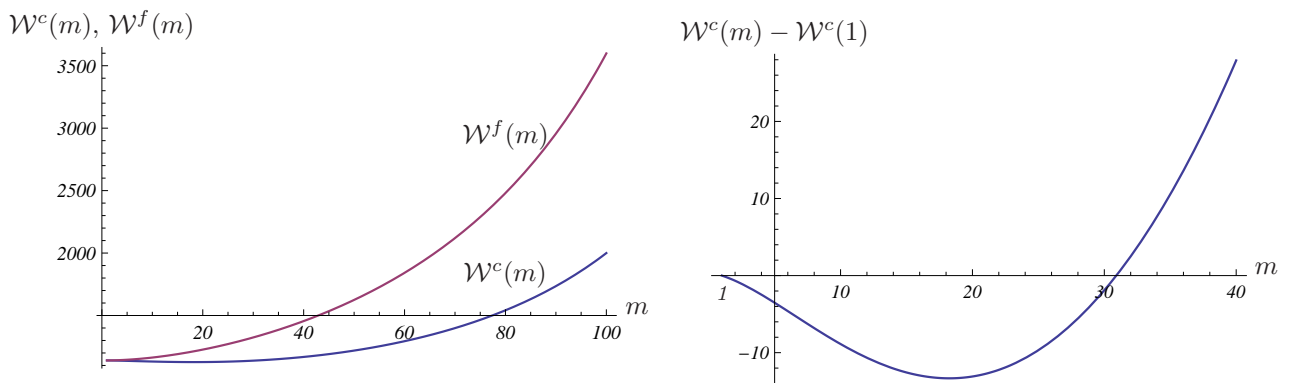


Figure 6: Welfare for coalition and fringe countries (Example 2)

The result $\mathcal{W}_m^c(1) < 0$ from Proposition 5(ii) gives rise to the question as to whether the initial dip of the coalition countries' welfare below their BAU welfare is always as small

²⁰Carraro and Siniscalco (1993, p. 313) denote such coalitions as *profitable*.

and negligible as in Example 1. To verify that coalitions of size $m \geq 2$ may be unprofitable we consider another parameter constellation, referred to as Example 2, that differs from Example 1 only in that $\alpha = 1000$ is replaced by $\alpha = 10000$. As illustrated in the right panel of Figure 6 the welfare difference $\mathcal{W}^c(m) - \mathcal{W}^c(1)$ is negative for all climate coalitions of size $m \in \{2, \dots, 30\}$ but positive and increasing for $m > 31$. The reason why the dip below zero of the difference $\mathcal{W}^c(m) - \mathcal{W}^c(1)$ is much more pronounced in Example 2 than in Example 1 is the increase in the parameter α because both examples differ only in that parameter. As suggested in footnote 9, α can be interpreted as a measure for the degree of progressivity of extraction costs. This result combining with the patterns we have observed in many other simulations not documented here leads us to put forward the following

Conjecture 1. *Suppose the functions T , V and D are specified as in (10) and consider $\tilde{m} > 1$ satisfying $\mathcal{W}^c(\tilde{m}) = \mathcal{W}^c(1)$ and $\mathcal{W}^c(m) < \mathcal{W}^c(\tilde{m})$ for all $m \in]1, \tilde{m}[$. \tilde{m} exists and*

- (i) is the larger, ceteris paribus, the more progressive are the extraction costs (the larger α) and the less declining is the marginal utility of the fuel (the smaller γ),*
- (ii) is independent of the level of the marginal utility of fuel (the level of β).*

4.3 Stability of climate coalitions

In the large literature on international environmental agreements and coalition formation²¹ the concept of stability (or equivalently, self-enforcement) is important because sovereign countries are expected to follow their self-interest (free-riding incentives) rather than committing to joint action and are therefore reluctant to accept a supranational authority for effective enforcement of international cooperation. The stability concept was first applied to international environmental agreements by Barrett (1994) and since then became the standard criterion for stable climate agreements. In the present context the coalition of some given size $m \in \{2, \dots, n\}$ is said to be stable, if $\mathcal{W}^c(m) \geq \mathcal{W}^f(m-1)$ (internal stability) and $\mathcal{W}^f(m) \geq \mathcal{W}^c(m+1)$ (external stability). Following Hoel and Schneider (1997) we define the stability function $\Phi(m) := \mathcal{W}^c(m) - \mathcal{W}^f(m-1)$ and readily conclude that the coalition of size m is stable if $\Phi(m) \geq 0$ (internal stability) and $\Phi(m+1) \leq 0$ (external stability). Obviously, the properties of the stability function Φ determine the existence, size and number of stable coalitions. Unfortunately, the function Φ is too complex to allow for deriving informative *analytical* results. We therefore proceed with calculating the stability functions of the Examples 1 and 2. We have plotted these functions in Figure 7 where we take the interval $[1, n]$ as the domain of Φ although the economically relevant domain is

²¹See e.g. Finus (2003) and the literature reviewed there.

$\{1, \dots, n\}$. Both stability functions start at $\Phi(1) = 0$. While in Example 1 ($\alpha = 1000$) the stability function is positive valued for $1 < m < 2.3$ and negative valued for $2.3 < m < 100$, in Example 2 ($\alpha = 10000$) $\Phi(m)$ is negative for all values of m . As an implication, stable coalitions (of size $m \geq 2$) do not exist in Example 2, but there is a unique stable coalition in Example 1 whose size is $m = 2$.

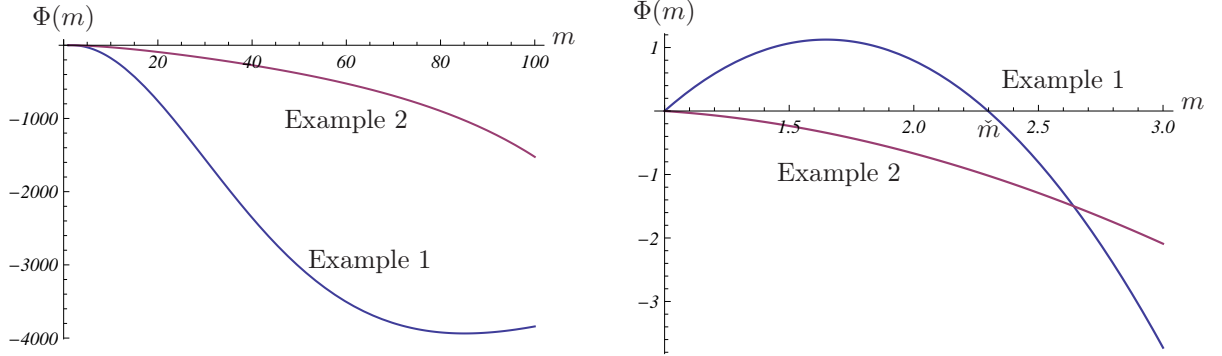


Figure 7: The stability functions $\Phi(m)$ for the Examples 1 and 2

We have conducted many more simulations with various other parameter constellations. The common feature of all of them is the result we have presented here for the Examples 1 and 2. Either there is a unique stable coalition consisting of two countries - as in Example 1 - or there exists no stable coalition at all - as in Example 2. The latter case appears to be the more likely, the larger is the interval $]1, \check{m}[$ defined in the Conjecture 1 above.²² We summarize our findings in

Conjecture 2. *If the functions T , V and D are specified as in (10), either no stable coalition exists or it consists of two countries.*

5 Is international trade harmful for the environment?

A straightforward way to understand the role of international trade for the formation and impact of climate coalitions is to compare our results with those of the autarky scenario in the otherwise unchanged model. The only substantive modification of the model (1) - (9) is to replace (3) by

$$x_i^s = x_i^d \quad \text{and} \quad e_i^s = e_i^d \quad i = 1, \dots, n \quad (21)$$

²²The role of the extraction costs will be further clarified in Section 5.

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 $i = 1, \dots, n$), but (5) is now replaced by $p_{ei} = -T'(e_i)$. With these modifications the welfare of country i is given by

$$W^i(e_1, \dots, e_n) = V(e_i) + T(e_i) - D\left(\sum_j e_j\right) \quad (22)$$

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

$$W^i(e_1, \dots, e_n) = \bar{x} + \beta e_i - \frac{\alpha + \gamma}{2} e_i^2 - \frac{\delta}{2} \left(\sum_j e_j\right)^2 \quad (23)$$

for the parametric functions (10). The basic model of coalition formation as discussed e.g. by Finus (2003) is characterized by a general functional form such as

$$W^i(e_1, \dots, e_n) = H(e_i) - D\left(\sum_j e_j\right) \quad (24)$$

with $H' > 0$ and $H'' < 0$ or by a parametric function

$$W^i(e_1, \dots, e_n) = \hat{\beta} e_i - \frac{\hat{\delta}}{2} e_i^2 - \frac{\hat{\rho}}{2} \left(\sum_j e_j\right)^2 \quad (25)$$

with $\hat{\beta}$, $\hat{\rho}$ and $\hat{\delta}$ being positive constants. Obviously, (22) and (24) (almost)²³ coincide, if we set $H(e_i) = V(e_i) + T(e_i)$, and (23) and (25) almost coincide, if $\hat{\beta} = \beta$, $\hat{\delta} = \delta$ and $\hat{\rho} = \alpha + \gamma$. In other words that basic model can be interpreted as the model of the present paper in the autarky regime (called 'autarky model', for short). Note also that in the autarky model the BAU equilibrium is determined by (15) and hence coincides with the BAU equilibrium under free trade. That result is not general, because it follows from the assumption of identical countries and the implied absence of international trade in the free-trade model.

In the following Section 5.1 we take coalition sizes as given and investigate the impact of trade liberalization, i.e. the impact of moving from autarky to free trade, on emissions and welfare. Section 5.2 takes up the issue of coalition stability again focusing on the question what the impact of trade liberalization is on the size of stable coalitions.

²³ $V' + T'$ may become negative for large e_i , but $H' > 0$ for all e_i . Yet $V' + T' < 0$ is an economically irrelevant case. If $\hat{\beta} = \beta$, $\hat{\delta} = \delta$ and $\hat{\rho} = \alpha + \gamma$ holds, (23) and (25) still differ by $\bar{x} > 0$. However, \bar{x} does not influence equilibria and hence can be ignored.

5.1 Trade liberalization, emissions and welfare

In case of autarky the coalition-fringe equilibrium $[me_c^a, (n - m)e_f^a]$ is characterized by the first-order conditions²⁴

$$V'(e_f^a) = -T'(e_f^a) + D'(e^a), \quad (26)$$

$$V'(e_c^a) = -T'(e_c^a) + mD'(e^a). \quad (27)$$

(26) and (27) yield negatively sloped best-reply curves. Since the reaction functions and first-order conditions under autarky exhibit the same quantitative properties than those under free trade (18) and (19), the Propositions 1 - 5 and the Figures 2 - 4 also apply in the autarky regime. The results of the Propositions 1 - 5 for the autarky regime are well known for linear and quadratic functional forms of the functions $V + T$ and D from the literature (Carraro and Siniscalco 1991, Finus 2001, De Cara and Rotillon 2001, Finus and Rundshagen 2001 and Rubio and Casino 2001). The contribution of our paper is to have shown that for *general* functions V, T and D the Propositions 1 - 5 hold in the autarky regime as well as in the free-trade regime.

Next, we investigate how opening the borders affects the emissions in the coalition-fringe equilibrium. Comparing those equilibrium emissions under autarky and free trade, the Appendix D proves

Proposition 6. *Let the functions T, V and D be specified as in (10) and consider the coalition-fringe equilibria in autarky, $[me_c^a, (n - m)e_f^a]$, and in the free-trade regime, $[me_c^*, (n - m)e_f^*]$, for some given coalition size m . Then the ranking of equilibrium emissions is $e_c^* < e_c^a < e_f^a < e_f^*$ and $e^* > e^a$.*

The ranking of emissions in Proposition 6 is an interesting piece of information. Trade liberalization, i.e. the transition from autarky to free trade, exacerbates climate damage ($e^* > e^a$). Although the coalition countries reduce their own emissions ($e_c^* < e_c^a$), all fringe countries expand their emissions ($e_f^* > e_f^a$) when moving from autarky to free trade to such an extent that the coalition's aggregate emission reduction is overcompensated. In the language of carbon leakage we find that trade liberalization induces carbon leakage at a rate greater than 100 %.

Proposition 6 is illustrated in Figure 3. As already noted in the discussion of Figure 3 in Section 4.1, the line AB is the graph of the fringe group's best-reply function R^f from (15) in the free-trade regime. The line KL in Figure 3 (which we had ignored in Section 4.1) is the graph of group F 's best-reply function in the autarky regime. The point of intersection, E_1 ,

²⁴In the sequel the autarky regime is indicated by the superscript a .

of the curves AB and KL characterizes a situation of zero trade flows in the free-trade regime implying that this intersection point represents the BAU equilibrium in both regimes. The line KL must be necessarily flatter than AB because otherwise it would not be consistent with the ranking of Proposition 6. A point such as E_3 in Figure 3 is that point on the curve KL , where the graph of the coalition's best-reply function in autarky intersects the curve KL . In line with the Propositions 2 and 6 we have positioned the equilibrium points E_2 and E_3 such that $e_c^* < e_c^a < e_o$ and $e_o < e_f^a < e_f^*$. Note also that Figure 3 displays three dotted straight lines each of which passes through one of the equilibrium points. By construction they have a negative slope of 45° such that world emissions e are constant on each of them and the level of world emissions increases with the distance of the line from the origin. Thus the dotted lines illustrate the ranking $e^a < e^* < ne_o$.

To further improve the intuition for the emission ranking of Proposition 6 we rewrite the first-order conditions characterizing the coalition-fringe equilibrium (18) and (19) under free trade as

$$V'(e_f^*) = \underbrace{\frac{\partial p_e}{\partial e_i} \cdot \frac{m}{n}(e_f^* - e_c^*)}_{>0} - T' \left(\frac{e^*}{n} \right) + D'(ne^*), \quad (28)$$

$$V'(e_c^*) + \underbrace{\frac{\partial p_e}{\partial e_i} \cdot \frac{m(n-m)}{n}(e_f^* - e_c^*)}_{>0} = -T' \left(\frac{e^*}{n} \right) + mD'(ne^*), \quad (29)$$

where $\frac{\partial p_e}{\partial e_i} \equiv -\frac{1}{n}T'' \left(\frac{e^*}{n} \right)$. The comparison of (28) and (29) with (26) and (27) reveals that opening borders changes the marginal costs reflected by the term $-T' \left(\frac{e}{n} \right)$ and causes additional marginal benefits or costs, captured by the term $\frac{\partial p_e}{\partial e_i} \cdot (e_f - e_c)$. Recall that in the autarky equilibrium the economy is characterized by $e_c^a < \frac{e^a}{n} < e_f^a$ (see Proposition 2 applied to the autarky regime) and suppose for a moment that borders are opened while the countries retain their emission caps e_c^a and e_f^a . Due to $-T'(e_f^a) > -T' \left(\frac{e^a}{n} \right) > -T'(e_c^a)$, trade liberalization reduces the marginal costs of coalition countries, ceteris paribus, and raises the marginal costs of fringe countries. Next, we turn to those marginal benefits or costs which result from strategic manipulation of the fuel price p_e in the free-trade regime. When borders are opened and either a coalition country or a fringe country relaxes its cap marginally, then the demand for fuel increases, ceteris paribus, and thus boosts the fuel price ($\partial p_e / \partial e_i > 0$). Since coalition countries export fuel, the price hike increases their revenues from fuel exports. In contrast, the price hike adds to the fringe countries' costs, because it raises their fuel import bill. To sum up, if coalition countries relax their cap, they enjoy additional "strategic" benefits and reduced marginal costs and if fringe countries relax their cap, they suffer from additional "strategic" costs and increased marginal costs. Since countries balance marginal benefits and marginal costs, trade liberalization induces the coalition to tighten its cap and the fringe countries to relax their caps.

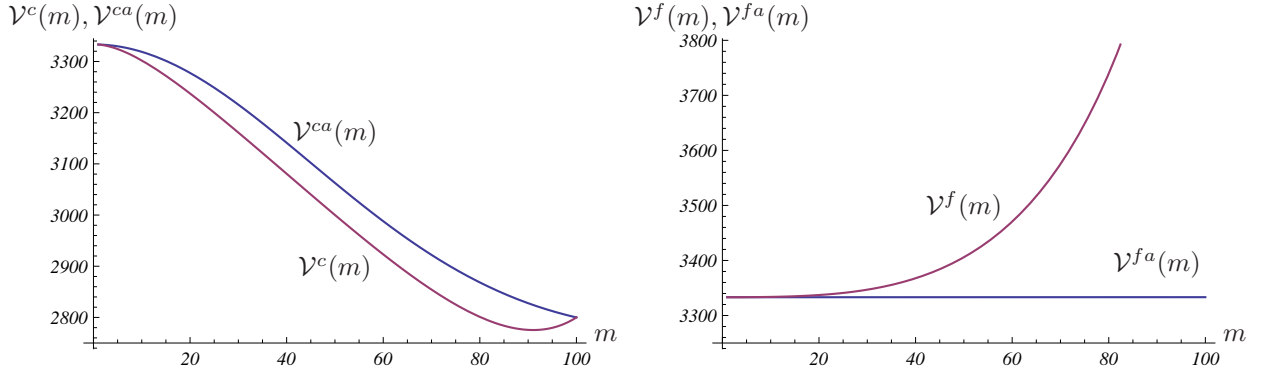


Figure 8: Autarky vs. free trade: non-climate welfare (Example 2)

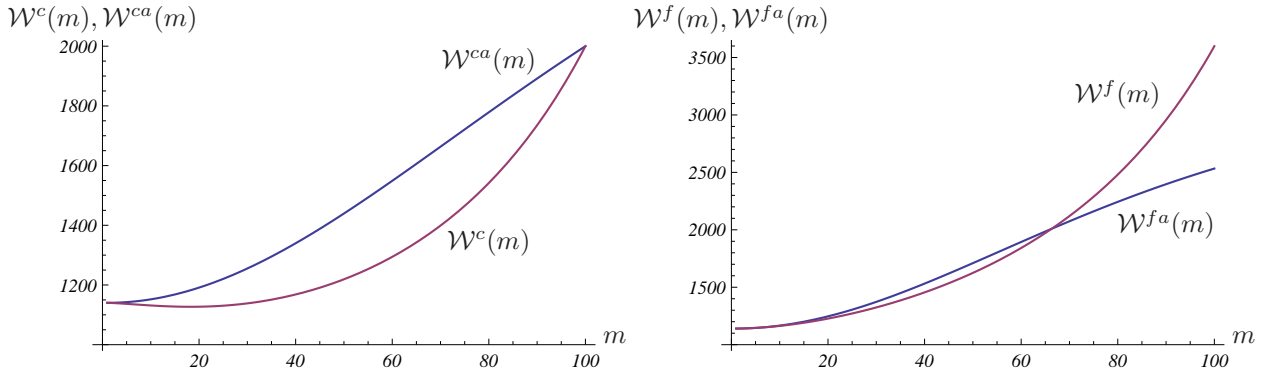


Figure 9: Autarky vs. free trade: total welfare (Example 2)

Opening the borders has significant welfare implications for coalition and fringe countries. It is clear that for all countries the climate welfare declines when moving from autarky to free trade, because international trade raises total emissions according to Proposition 6. Turning to non-welfare climate, Proposition 6 also reveals that fringe countries benefit from the opening of borders through increased consumption of fuel ($e_f^* > e_f^a$) while coalition countries suffer from less fuel consumption ($e_c^* < e_c^a$). However, recall from (11) that non-climate welfare not only depends on fuel consumption but also on world market interdependence effects which are opposite to the fuel consumption effects. To be more specific, the coalition country's strategic effect $-\frac{n-m}{n}(e_f^* - e_c^*)T'(\frac{e^*}{n}) > 0$ enhances non-climate welfare, whereas the fringe country's strategic effect $-\frac{m}{n}(e_c^* - e_f^*)T'(\frac{e^*}{n}) < 0$ reduces non-climate welfare.

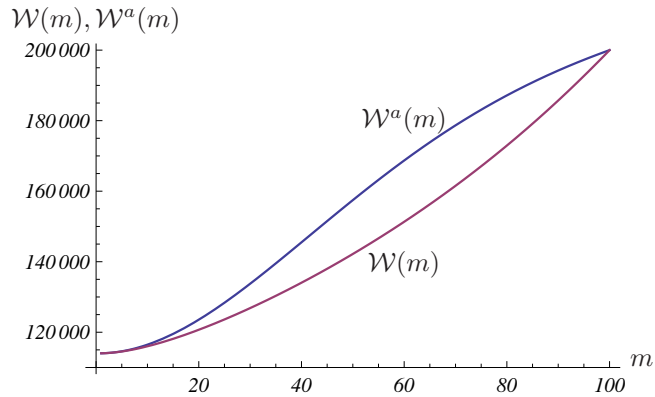


Figure 10: Autarky vs. free trade: Aggregate welfare of all countries (Example 2)

Since the complexity of the effects does not allow for clear-cut analytical results concerning non-climate welfare and total welfare, we resort to the Figures 8 through 10 that are based on the Example 2. According to Figure 8 the non-climate welfare of fringe [coalition] countries denoted by $\mathcal{V}^f(m)$ [$\mathcal{V}^c(m)$] is increasing [decreasing] in the transition from autarky to trade. Hence, we conclude that the fuel consumption effects overcompensate the world-market interdependence effects. Since both non-climate welfare and climate welfare decrease for coalition countries when the borders are opened, the coalition countries suffer a loss of total welfare from trade liberalization as shown in the left panel of Figure 9. In contrast, the fringe countries' changes of non-climate welfare and climate welfare are opposite in sign. The right panel of Figure 9 demonstrates that opening the borders is welfare-enhancing [welfare reducing] for large [small] coalition-sizes m . Finally, Figure 10 displays the aggregate welfare of all countries, $\mathcal{W}(m) := m\mathcal{W}^c(m) + (n - m)\mathcal{W}^f(m)$. Trade liberalization is not only harmful for the environment but also reduces aggregate welfare. The welfare loss through climate damage exceeds the potential non-climate welfare gain. We found that these results also hold for a large set of other parameter constellations and summarize them in

Conjecture 3. *Trade liberalization leads to the following welfare changes:*

- (i) *The climate welfare decreases in all countries.*
- (ii) *The non-climate welfare declines in coalition countries and rises in fringe countries.*
- (iii) *The total welfare of coalition countries decreases.*
- (iv) *The total welfare of fringe countries increases (decreases) for large (small) coalition sizes.*
- (v) *The aggregate welfare of all countries decreases.*

5.2 Trade liberalization and stability of coalitions

Finally, we turn to the impact of trade liberalization on the size of stable coalitions. Finus and Rundshagen (2001), Rubio and Casino (2001) and de Cara and Rotillon (2001) have independently proved that the self-enforcing IEA consists of at most two countries in the quadratic basic model of the literature, (25). From (23) and (25) follows that if a stable coalition of size $m = 2$ exists [does not exist] in the economy defined by the parameters $(\hat{\beta}, \hat{\delta}, \hat{\rho})$ in (25), it also exists [does not exist] in every economy $(\alpha, \beta, \gamma, \delta) \in E(\hat{\beta}, \hat{\delta}, \hat{\rho})$, where $E(\hat{\beta}, \hat{\delta}, \hat{\rho}) := \left\{ (\alpha, \beta, \gamma, \delta) \in \mathbb{R}_+^4 \mid \beta = \hat{\beta}, \delta = \hat{\delta} \text{ and } \alpha + \gamma = \hat{\rho} \right\}$. That observation gives rise to the interesting question, whether or not the pattern of coalition stability changes from a two-country stable coalition to no stable coalition or vice versa, when an economy in $E(\hat{\beta}, \hat{\delta}, \hat{\rho})$ switches from the autarky regime to the free-trade regime.

| Economy | $\beta = \hat{\beta}$ | $\delta = \hat{\delta}$ | α | γ | $\alpha + \gamma = \hat{\rho}$ |
|---------|-----------------------|-------------------------|----------|----------|--------------------------------|
| E_1 | 10000 | 1 | 0 | 15000 | 15000 |
| E_2 | 10000 | 1 | 1500 | 13500 | 15000 |
| E_3 | 10000 | 1 | 5000 | 10000 | 15000 |
| E_4 | 10000 | 1 | 10000 | 5000 | 15000 |

Table 1: The economies $E_1 - E_4$

To answer that question, we consider the basic model with parameters $(\hat{\beta} = 10000, \hat{\delta} = 1, \hat{\rho} = 15000)$ and define the economies E_1 through E_4 as specified in Table 1. By construction, these four economies belong to $E(\hat{\beta}, \hat{\delta}, \hat{\rho})$ and hence are equivalent to the 'basic economy' in the autarky regime. The economy E_1 is a polar case that we can use to represent the economy $(\hat{\beta}, \hat{\delta}, \hat{\rho})$, because (10), $\alpha = 0$ and $\gamma = 15000$ turn (11) into (25). $\alpha = 0$ means zero extraction costs in terms of good X . Hence there is no international trade in E_1 even if the borders are open. The curve $\Phi^a(m)$ in Figure 11 is the stability curve²⁵ for E_1 through E_4 in autarky. Its shape implies that there exists a stable coalition of size $m = 2$ in the basic economy $(\hat{\beta}, \hat{\delta}, \hat{\rho})$ as well as in the autarky regime of the economies E_1 through E_4 . The crucial question is whether that stable coalition prevails in the economies E_2, E_3 and E_4 , when we change the regime from autarky to free trade.

In Figure 11 the curve marked by $\Phi(m; E_i)$ represents the stability curve in the free-trade regime of the economy $E_i, i = 2, 3, 4$. According to the shape of these stability curves a two-country stable coalition still exists for economy E_2 , but not any more for the economies

²⁵We have calculated the stability function $\Phi^a(m)$ by first determining the stability function of one of the economies E_2, E_3 or E_4 in the free-trade regime and then replacing the parameters α and γ by $\alpha = 0$ and $\gamma = 15000$.

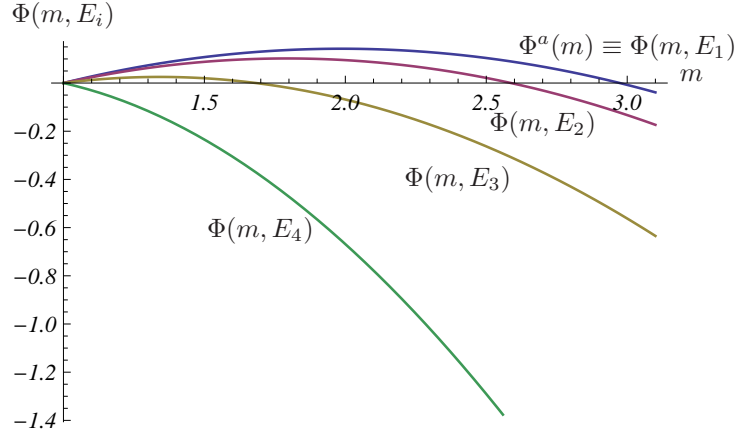


Figure 11: Stability functions of the economies $E_1 - E_4$

E_3 and E_4 . The larger the parameter α is at the expense of the parameter γ the less likely it becomes to form the stable two-country coalition. While the supply parameter α (which reflects the magnitude and progressivity of extraction costs) and the demand parameter γ (which reflects the diminishing marginal utility of fuel) are perfect substitutes in autarky, their impact differs in the free-trade regime. The key for understanding those differential effects is the Conjecture 1, that establishes the link between variations of the parameters α and γ on the one hand and the profitability of coalitions in the case of international trade on the other hand. Increasing α and decreasing γ render the coalition less profitable and shift the stability curve downward and to the left as shown in Figure 11. These findings suggest

Conjecture 4. *If the functions T, V and D are specified as in (10), then trade liberalization tends to render more unlikely the formation of a two-country stable coalition.*

6 Concluding remarks

In the present paper we study a stylized model of the world economy in which fossil fuel and a composite consumer good are produced in symmetric countries and traded on world markets. Carbon emissions from fuel consumption create global climate damage. If world markets do not exist (autarky regime) the model turns out to coincide with the *basic model* of standard theory of climate coalition formation. That gives rise to the question what the impact of trade liberalization, i.e. the move from autarky to free trade, is (i) for the performance of given coalitions and their fringe and (ii) for the formation of self-enforcing IEAs. The principal reason for expecting differences in outcome is that in the free-trade regime domestic

mitigation policies not only have an impact on climate change but also on the relative world fuel price (terms-of-trade) which coalitions and countries seek to manipulate in their own favor.

We find that in the quadratic version of our free-trade model stable coalitions consist of at most two countries like in the autarky model. That is, the terms-of-trade effects turn out to be second-order effects. Insofar, extending the basic model by international trade does not modify the outcome. We did not see clear reasons for that result ex ante and it is therefore not an obvious result. The same (maximum) number of countries in stable coalitions in free trade and autarky does not mean that moving from autarky to trade is inconsequential. On the contrary, trade liberalization sets off a number of remarkable changes. In particular, world emissions rise although the coalition steps up its mitigation effort ('green paradox of trade liberalization'). Trade liberalization also turns out to be bad for the coalition countries' welfare and the aggregate welfare of all countries; it reduces the range of profitable coalitions, and it even tends to hamper the formation of stable coalitions.

Although our model has more 'economic' structure than the basic model, we have kept it simple for the benefit of comparing it with the basic model and of deriving informative results. Severely restrictive assumptions are a high price to be paid for analytical tractability that we share with the extant theoretical literature on climate coalition formation. One would want to check the robustness of results when economies are modeled in a more complex way which would be an assignment for computable large-scale models. For future analytical approaches various routes - always combined with international trade - are desirable and some are also tractable as we believe. One such step toward a more realistic setting would be to model fossil fuel not only as a final consumption good but also as an intermediary industrial input. Similarly, while the assumption of symmetric countries is chosen to get for informative results, it abstracts from an important real-world complexity and likely underestimates the difficulties of forming stable coalitions. A modest but important step would be to consider two subsets of countries that differ with respect to the costs of fuel extraction.

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. (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 papers 568.
- Carraro, C. and D. Siniscalco (1993): Strategies for the international protection of the environment, *Journal of Public Economics* 52, 309-328.
- Copeland, B. and M.S. Taylor (1995): Trade and transboundary pollution, *American Economic Review* 85, 716-737.
- De Cara, S. and G. Rotillon (2001): Multi greenhouse gas international agreements, mimeo.
- 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): Self-enforcing environmental agreements (Stackelberg) and international trade, manuscript.
- Finus, M. (2001): *Game Theory and International Environmental Cooperation*, Cheltenham, Edward Elgar.
- Finus, M. and B. Rundshagen (2001): Endogenous coalition formation in global pollution control, working paper No. 43, 2001, Fondazione Eni Enrico Mattei, Milan.
- 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 Resource Economics* 2003/2004, Edward Elgar, Cheltenham.
- Fuentes-Albero, C. and S.J. Rubio (2010): Can international environmental cooperation be bought?, *European Journal of Operations Research* 216, 697-699.
- Hatzipanayatou, P., Lakiri, S. and M.S. Michael (2005): Globalization, cross-border pollution and welfare, CESifo working paper No. 1479.
- 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.
- Lapan, H.E. and S. Skidar (2011): Strategic environmental policy under free trade with transboundary pollution, *Review of Development Economics* 15, 1-18.
- Ludema, R. and J. Wooton (1994): Cross-border externalities and trade liberalization: The strategic control of pollution, *Canadian Journal of Economics* 27, 950-966.

- Markusen, J. (1975): International externalities and optimal tax structure, *Journal of International Economics* 5, 15-29.
- McGinty, M. (2007): International environmental agreements among asymmetric nations, *Oxford Economic Papers* 59, 45-62.
- Pavlova, Y. and A. de Zeeuw (2012): Asymmetries in international environmental agreements, manuscript.
- Rauscher, M. (1997): *International Trade, Factor Movements, and the Environment*, Oxford: Oxford University Press.
- Rauscher, M. (2001): On ecological dumping, in Batabyal, A. and M. Beladi (eds.): *The Economics of International Trade and the Environment*, Lewis Publishers.
- Rubio, S. and B. Casino (2001): International cooperation in pollution control, mimeo.
- Rubio, S. and A. Ulph (2006): Self-enforcing international environmental agreements revisited, *Oxford Economic Papers* 58, 233-263.
- Van der Pol, T., Weikard, H.-P. and E. van Ierland (2012): Can altruism stabilise international climate agreements?, *Ecological Economics*, in press.

Appendix

A. Country i 's welfare as a function of all countries' emissions

According to (10) the welfare of country i is $W^i(e_1, \dots, e_n) = V(e_i) - G^i(e_1, \dots, e_n)$ where

$$G^i(e_1, \dots, e_n) := D \left(\sum_j e_j \right) - T \left(\frac{\sum_j e_j}{n} \right) + \left(\frac{\sum_j e_j}{n} - e_i \right) T' \left(\frac{\sum_j e_j}{n} \right). \quad (\text{A1})$$

Differentiation of (A1) yields

$$G_{e_i}^i = D' + \left(\sum_j e_j - ne_i \right) \frac{T''}{n^2} - T' = \frac{nD' - (n - \varepsilon_{xe})T' - e_i T''}{n}, \quad (\text{A2})$$

$$G_{e_k}^i = D' + \left(\sum_j e_j - ne_i \right) \frac{T''}{n^2} \quad i \neq k, \quad (\text{A3})$$

$$G_{e_i e_i}^i = \left(\sum_j e_j - ne_i \right) \frac{T'''}{n^3} + \frac{n^2 D'' - (2n - 1)T''}{n^2}, \quad (\text{A4})$$

$$G_{e_i e_k}^i = \left(\sum_j e_j - ne_i \right) \frac{T'''}{n^3} + \frac{n^2 D'' - (n - 1)T''}{n^2} = G_{e_i e_i}^i + \frac{T''}{n} \quad i \neq k, \quad (\text{A5})$$

where $\varepsilon_{xe} := \frac{e_i^s T''}{T'} > 0$.

Lemma 1.

$$(i) \quad G_{e_i}^i > 0 \iff \varepsilon_{xe} < n + \frac{e_i T'' - n D'}{T'}, \text{ where } n + \frac{e_i T'' - n D'}{T'} > n > 0.$$

$$(ii) \quad G_{e_k e_k}^i > 0, \text{ if } T''' \in] - \bar{\tau}, \bar{\tau}[, \text{ where } \bar{\tau} := \frac{\bar{\phi}}{\bar{e}}, \bar{e} := n T^{-1}(x_i^s), E := \{(e_1, \dots, e_n) | e_i \in [0, \frac{\bar{e}}{n}]\}, \\ \bar{\phi} := \min_{(e_1, \dots, e_n) \in E} \left[(n^2 - n) T'' \left(\frac{\sum_j e_j}{n} \right) - n^3 D'' \left(\sum_j e_j \right) \right].$$

Proof of (ii). Observe first that (A5) implies

$$G_{e_i e_k}^i > 0 \iff \left(\sum_j e_j - n e_i \right) T''' > (n^2 - n) T'' - n^3 D''. \quad (\text{A6})$$

For $\left(\sum_j e_j - n e_i \right) = 0$, $G_{e_i e_k}^i > 0$ is obvious. Suppose next $\left(\sum_j e_j - n e_i \right) \neq 0$ and observe that $\left(\sum_j e_j - n e_i \right) \in] - \bar{e}, \bar{e}[$ is a necessary equilibrium condition. If $\bar{e} > \left(\sum_j e_j - n e_i \right) > 0$, then $-\bar{\tau} > \frac{(n^2 - n) T'' - n^3 D''}{\sum_j e_j - n e_i}$ and therefore $G_{e_i e_k}^i > 0$ follows from (A6), if $T''' > -\bar{\tau}$. Likewise, if $0 > \left(\sum_j e_j - n e_i \right) > -\bar{e}$, then $\bar{\tau} < -\frac{(n^2 - n) T'' - n^3 D''}{\sum_j e_j - n e_i}$ and therefore $G_{e_i e_k}^i > 0$ follows from (A6), if $T''' < \bar{\tau}$. Thus $T''' \in] - \bar{\tau}, \bar{\tau}[$ implies $G_{e_i e_k}^i > 0$ in the relevant domain of (e_1, \dots, e_n) . ■

Throughout the paper we assume the welfare function W^i to be strictly concave [the function G^i to be strictly convex in e_i] with an interior maximum in the relevant part of the domain. Sufficient conditions for that property are: $\varepsilon_{xe} < n + \frac{e_i T'' - n D'}{T'}$, $T''' \in] - \bar{\tau}, \bar{\tau}[$ and $V'(e_i) = \infty$ for $e_i = 0$.

B. Derivation of the best-reply functions (15) and their properties

We take $e_i = \tilde{R} \left(\sum_{j \neq i} e_j \right)$, $i = 1, \dots, n$, with $\tilde{R}_{\sigma_i} \in] - 1, 0[$ and $\sigma_i := \sum_{j \neq i} e_j$ from (13) as our point of departure. Set $e_i = e_c$ for all $i \in C$ and $e_i = e_f$ for all $i \in F$ to turn (13) into

$$e_f = \tilde{R}^f \left[\underbrace{m e_c + \frac{n - m - 1}{n - m} (n - m) e_f}_{=:\sigma_f} \right] \text{ for } i \in F \\ \wedge e_c = \tilde{R}^c \left[\underbrace{\frac{m - 1}{m} m e_c + (n - m) e_f}_{=:\sigma_c} \right] \text{ for } i \in C,$$

where $\tilde{R}^c(\sigma_c) = \tilde{R}^f(\sigma_f) = \tilde{R}(\sigma_i)$ if and only if $\sigma_c = \sigma_f = \sigma_i$. Summation over all countries in group F and in group C , respectively, yields

$$(n - m) e_f = (n - m) \tilde{R}^f \left[m e_c + \frac{n - m - 1}{n - m} (n - m) e_f \right] \\ \wedge m e_c = \tilde{R}^c \left[\frac{m - 1}{m} m e_c + (n - m) e_f \right]. \quad (\text{B1})$$

The equations (B1) can be conceived of as mappings from $(m e_c, m)$ into $(n - m) e_f$ and from $[(n - m) e_f, m]$ into $m e_c$, respectively.²⁶ In fact, they implicitly represent best-reply functions of the

²⁶Here we treat m , $m e_c$ and $(n - m) e_f$ as three different variables. m is a variable of its own as far as m is not part of $m e_c$ or $(n - m) e_f$.

type (15) satisfying $R_{me_c}^f \in]-1, 0[$ and $R_{(n-m)e_f}^c \in]-1, 0[$, because differentiation of (B1) yields

$$\begin{aligned} d(n-m)e_f &= (n-m)\tilde{R}_{\sigma_f}^f dme_c + (n-m-1)\tilde{R}_{\sigma_f}^f d(n-m)e_f, \\ dme_c &= (m-1)\tilde{R}_{\sigma_c}^c dme_c + m\tilde{R}_{\sigma_c}^c d(n-m)e_f \end{aligned}$$

or equivalently

$$\begin{aligned} \frac{d(n-m)e_f}{dme_c} &= R_{me_c}^f = -\frac{(n-m)\tilde{R}_{\sigma_f}^f}{(n-m)\tilde{R}_{\sigma_f}^f - (1 + \tilde{R}_{\sigma_f}^f)} \in]-1, 0[, \\ \frac{dme_c}{d(n-m)e_f} &= R_{(n-m)e_f}^c = -\frac{m\tilde{R}_{\sigma_c}^c}{m\tilde{R}_{\sigma_c}^c - (1 + \tilde{R}_{\sigma_c}^c)} \in]-1, 0[. \end{aligned}$$

C. Proof of Proposition 3

To prove $\mathcal{E}_m^f = -\mathcal{E}_m$, where $\mathcal{E}(m) := \mathcal{E}^c(m) + (n-m)\mathcal{E}^f(m)$, consider an individual fringe country's equilibrium welfare²⁷

$$W^f[\mathcal{E}^f(m), S^f(m)] = V[\mathcal{E}^f(m)] - G^f[\mathcal{E}^f(m), S^f(m)] \quad (C1)$$

with $\sigma_f = S^f(m) = m\mathcal{E}^c(m) + (n-m-1)\mathcal{E}^f(m) = \mathcal{E}(m) - \mathcal{E}^f(m)$. The first-order condition of maximizing W^f with respect to e_f is $V'(e_f) - G_{e_f}^f(e_f, \sigma_f) = 0$. Total differentiation yields $V''de_f - G_{e_f e_f}^f de_f - G_{e_f \sigma_f}^f d\sigma_f$ and $(V'' - G_{e_f e_f}^f)de_f = \frac{nG_{e_f e_f}^f + T''}{n}d\sigma_f$ because $G_{e_f \sigma_f}^f = G_{e_f e_f}^f + \frac{T''}{n} > 0$ (which follows from differentiation of (A3)). That leads to

$$\frac{de_f}{d\sigma_f} = -\frac{nG_{e_f e_f}^f + T''}{nG_{e_f e_f}^f - nV''} \in]-1, 0[. \quad (C2)$$

Combine (C2) with $\frac{d\sigma_f}{dm} = \frac{de}{dm} - \frac{de_f}{dm}$ and $\frac{de_f}{dm} = \frac{de_f}{d\sigma_f} \cdot \frac{d\sigma_f}{dm}$ to obtain

$$\frac{de_f}{dm} = -\underbrace{\frac{n(V'' - G_{e_f e_f}^f)}{nV'' + T''}}_+ \cdot \frac{de}{dm} \implies \text{sign } \mathcal{E}_m^f = -\text{sign } \mathcal{E}_m. \quad (C3)$$

Next consider an individual country, say country ξ , that belongs to the coalition before the coalition size increases from m_o to $m_o + 1$. The pertaining first-order conditions are

$$\begin{aligned} V'[\mathcal{E}^\xi(m_o)] - m_o G_{me_c}^c[\mathcal{E}^\xi(m_o), \mathcal{E}(m_o)] &= 0, \\ V'[\mathcal{E}^\xi(m_o + 1)] - (m_o + 1)G_{me_c}^c[\mathcal{E}^\xi(m_o + 1), \mathcal{E}(m_o + 1)] &= 0, \end{aligned}$$

where

$$G_{me_c}^c[\mathcal{E}^\xi(m), \mathcal{E}(m)] = D'[\mathcal{E}(m)] + \frac{\mathcal{E}(m) - n\mathcal{E}^\xi(m)}{n^2} \cdot T'' \left[\frac{\mathcal{E}(m)}{n} \right] - \frac{-T' \left[\frac{\mathcal{E}(m)}{n} \right]}{m} =: \hat{G}_{me_c}^c(m).$$

²⁷To ease the notation we drop the argument m in $\mathcal{E}^c(m)$ etc. if there is no risk of confusion.

For later use define the function

$$A(m) := D'[\mathcal{E}(m)] + \frac{\mathcal{E}(m) - n\mathcal{E}^\xi(m)}{n^2} \cdot T'' \left[\frac{\mathcal{E}(m)}{n} \right] - T' \left[\frac{\mathcal{E}(m)}{n} \right].$$

Observe that $\hat{G}_{me_c} = A(m) - \frac{m-1}{m}T'[\mathcal{E}(m)]$. Since $V'' < 0$, we get

$$\begin{aligned} \mathcal{E}^\xi(m_o) \geq \mathcal{E}^\xi(m_o + 1) &\iff m_o \hat{G}_{me_c}^c(m_o) \leq \hat{G}_{me_c}^c(m_o + 1) \\ \iff m_o A(m_o) - (m_o - 1)T'[\mathcal{E}(m_o)] &\leq (m_o + 1)A(m_o + 1) - m_o T'[\mathcal{E}(m_o + 1)]. \end{aligned} \quad (C4)$$

Differentiation of $A(m)$ with respect to m yields

$$\frac{dA}{dm} = \frac{\partial A}{\partial e_\xi} \mathcal{E}_m^\xi + \frac{\partial A}{\partial e} \mathcal{E}_m, \quad (C5)$$

where $\frac{\partial A}{\partial e_\xi} = -\frac{T''}{n} > 0$ and $\frac{\partial A}{\partial e} = D'' - \frac{(n-1)T''}{n^2} + \frac{(e-ne_c)T'''}{n^3} = G_{e_i e_k}^i > 0$ (with $G_{e_i e_k}^i > 0$ shown in Appendix A). Contrary to the assertion suppose that $\mathcal{E}_m > 0$. Then $\mathcal{E}_m^f < 0$ due to (C3), and therefore $\mathcal{E}_m^c > 0$ is a necessary condition for

$$\mathcal{E}_m = e_c - e_f + m\mathcal{E}_m^c + (n - m)\mathcal{E}_m^f \quad (C6)$$

to be positive. Accounting for (C5) and (C6) we find

$$\begin{aligned} (\mathcal{E}_m > 0 \quad \wedge \quad \mathcal{E}_m^c > 0) &\implies \frac{dA}{dm} > 0 \implies A(m_o + 1) > A(m_o) \\ &\implies (m_o + 1)A(m_o + 1) > m_o A(m_o). \end{aligned} \quad (C7)$$

In addition, due to $T' < 0, T'' < 0$ it holds that

$$\begin{aligned} \mathcal{E}_m > 0 &\implies -T'[\mathcal{E}(m_o)] < -T'[\mathcal{E}(m_o + 1)] \\ &\implies -(m_o - 1)T'[\mathcal{E}(m_o)] < -m_o T'[\mathcal{E}(m_o + 1)]. \end{aligned} \quad (C8)$$

From (C4), (C7) and (C8) we infer

$$[\mathcal{E}_m > 0 \quad \wedge \quad \mathcal{E}_m^c > 0] \implies [\mathcal{E}^\xi(m_o + 1) < \mathcal{E}^\xi(m_o)] \implies \mathcal{E}_m^c < 0.$$

That contradiction proves $\mathcal{E}_m < 0$, which in turn implies $\mathcal{E}_m^f > 0$ via (C3). ■

D. Proof of Proposition 5

(i) Consider an individual fringe country's welfare (C1). Differentiation with respect to m yields

$$\begin{aligned} W_m^f &= V' \mathcal{E}_m^f - G_{e_f}^f \mathcal{E}_m^f - G_{\sigma_f}^f S_m^f \\ &= \underbrace{(V' - G_{e_f}^f)}_{=0} \mathcal{E}_m^f - G_{\sigma_f}^f S_m^f = -G_{\sigma_f}^f S_m^f. \end{aligned} \quad (D1)$$

$W_m^f > 0$ because $S_m^f = \mathcal{E}_m - \mathcal{E}_m^f < 0$ follows from Proposition 3 and $G_{\sigma_f}^f = G_{e_k}^i > 0$ ($i \neq k$) follows from Appendix A.

(ii) Consider an individual coalition country's equilibrium welfare

$$W^c[\mathcal{E}^c(m), S^c(m)] = V[\mathcal{E}^c(m)] - G^c[\mathcal{E}^c(m), S^c(m)] \quad (\text{D2})$$

with $\sigma_c = S^c(m) = (m-1)\mathcal{E}^c(m) + (n-m)\mathcal{E}^f(m)$. Differentiation of (D2) with respect to m yields

$$W_m^c(m) = V'\mathcal{E}_m^c - G_{e_c}^c\mathcal{E}_m^c - G_{\sigma_c}^c S_m^c. \quad (\text{D3})$$

Observe that it holds $V' - G_{e_c}^c = 0$ at $m = 1$. Hence we have $W_m^c(1) = -G_{\sigma_c}^c S_m^c(1)$. Next, verify that $G_{\sigma_c}^c = G_{e_k}^i > 0$ ($i \neq k$) from Appendix A and that differentiation of $S^c(m)$ yields

$$S_m^c(m) = \mathcal{E}^c(m) - \mathcal{E}^f(m) + (m-1)\mathcal{E}_m^c + (n-m)\mathcal{E}_m^f. \quad (\text{D4})$$

Due to $\mathcal{E}^c(1) = \mathcal{E}^f(1)$ the equation (D4) evaluated at $m = 1$ turns into $S_m^c(1) = (n-m)\mathcal{E}_m^f$. Combined with $G_{\sigma_c}^c > 0$ and $\mathcal{E}_m^f > 0$ (see Proposition 3) we get $W_m^c(1) < 0$. ■

E. Proof of Proposition 6

Consider first the case of international trade. Using the parametric function (10) in the first-order conditions (18) and (19) yields, after some rearrangement of terms,

$$e_c = \frac{\beta n^2}{\alpha m(2n-m) + (\delta m^2 + \gamma)n^2} - \frac{\alpha(n-m) + \delta n^2 m}{\alpha m(2n-m) + (\delta m^2 + \gamma)n^2} e_f, \quad (\text{E1})$$

$$e_f = \frac{\beta n^2}{\alpha(n^2 - m(n-1)) + (\delta(n-m) + \gamma)n^2} - \frac{\alpha(n-1) + \delta n^2}{\alpha(n^2 - m(n-1)) + (\delta(n-m) + \gamma)n^2} e_c \quad (\text{E2})$$

Solving (E1) and (E2) with respect to e_c and e_f we obtain the caps

$$e_c^* = \frac{\beta[n^2\gamma + \alpha m(n+1-m) + n^2\delta(n-m)(m-1)]}{B} \quad (\text{E3})$$

$$e_f^* = \frac{\beta[n^2\gamma + \alpha m(n+1-m) + n^2\delta m(m-1)]}{B} \quad (\text{E4})$$

$$m e_c^* + (n-m)e_f^* = \frac{\beta n[n^2\gamma + \alpha m(n+1-m) + n^2]}{B}, \quad (\text{E5})$$

in the free-trade coalition-fringe Nash equilibrium $[m e_c^*, (n-m)e_f^*]$, where

$$B := n^2\gamma(\alpha + \gamma + n\delta) + m^2(n^2\gamma\delta - \alpha\gamma - \alpha^2) + m[\alpha(\gamma + n\gamma + n^2\delta) + (1+n)\alpha^2 - n^2\gamma\delta] > 0.$$

In case of autarky the coalition-fringe Nash equilibrium $[m e_c^a, (n-m)e_f^a]$ is characterized by

$$e_c^a = \frac{\beta[\alpha + \gamma - (m-1)(m-n)\delta]}{(\alpha + \gamma)[\alpha + \gamma + \delta(m(m-1) + n)]}, \quad (\text{E6})$$

$$e_f^a = \frac{\beta[\alpha + \gamma + m(m-1)\delta]}{(\alpha + \gamma)[\alpha + \gamma + \delta(m(m-1) + n)]}, \quad (\text{E7})$$

$$m e_c^a + (n-m)e_f^a = \frac{n\beta}{(\alpha + \gamma)[\alpha + \gamma + \delta(m(m-1) + n)]}. \quad (\text{E8})$$

A comparison of (E3) - (E5) and (E6) - (E8) yields after tedious rearrangement of terms

$$e_c^* - e_c^a = -\frac{(m-1)(n-m)n\alpha\beta\delta[m^2n\delta + n(\alpha + \gamma + n\delta) - m(\alpha + \gamma + 2n\delta)]}{B(\alpha + \gamma)[\alpha + \gamma + \delta(m(m-1) + n)]} < 0, \quad (\text{E9})$$

$$e_f^* - e_f^a = \frac{(m-1)mna\beta\delta[(n-1)(\alpha + \gamma) + n\delta(n + 2m^2 - 2m)]}{B(\alpha + \gamma)[\alpha + \gamma + \delta(m(m-1) + n)]} > 0 \quad (\text{E10})$$

and

$$me_c^* + (n-m)e_f^* - me_c^a - (n-m)e_f^a = \frac{(m-1)^2m(n-m)n\alpha\beta\delta}{B(\alpha + \gamma)[\alpha + \gamma + \delta(m(m-1) + n)]} > 0. \quad (\text{E11})$$

■

Appendix F. Autarky and international trade in Example 2: Equilibrium emissions (only for the referees)

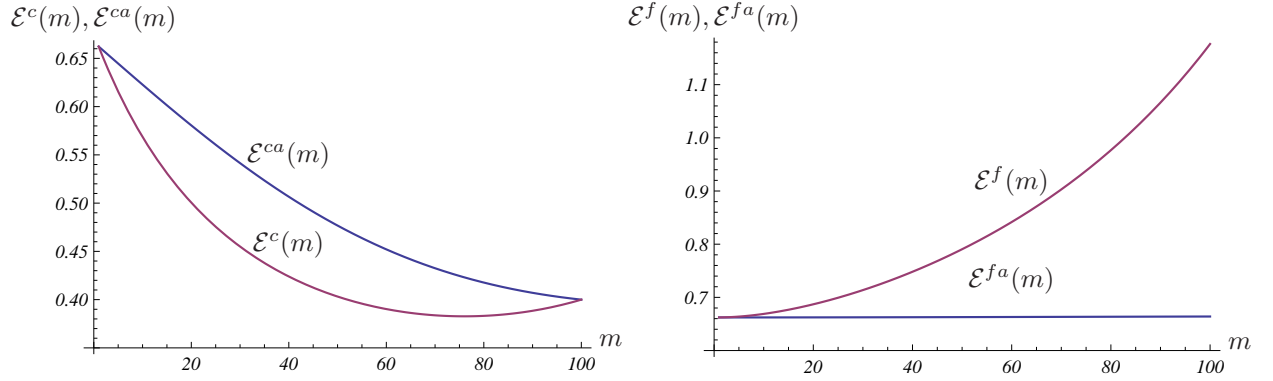


Figure 12: Equilibrium emission caps

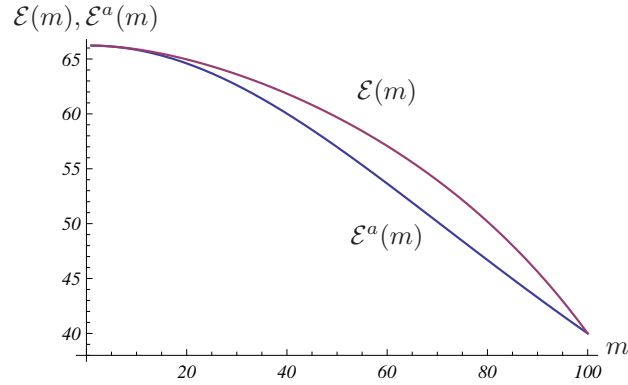


Figure 13: Total emissions