Climate cooperation with technology investments and border carbon adjustment

Carsten Helm* Robert C. Schmidt†

March 3, 2014

Abstract

A central question in climate policy is whether early investments in low-carbon technologies are a useful first step towards a more effective climate agreement in the future. We introduce a climate cooperation model with endogenous R&D investments where countries protect their international competitiveness via border carbon adjustments (BCA). BCA raises the scope for cooperation and leads to a non-trivial relation between countries' prior R&D investments and participation in the coalition. We find that early investments in R&D render free-riding more attractive. Therefore, with delayed cooperation on emission abatement and ex-ante R&D investments, the outcome is often characterized by high participation but inefficiently low technology investments and abatement.

*Carl von Ossietzky Universität Oldenburg, Department of Business Adminstration, Economics and Law, 26111 Oldenburg, Germany; E-mail: carsten.helm@uni-oldenburg.de

†Department of Economics, Humboldt University Berlin, Spandauer Str. 1, 10178 Berlin, Germany; E-mail: robert.schmidt.1@wiwi.hu-berlin.de

Keywords: climate treaty, border carbon adjustment, border tax adjustment, coalitions, R&D

JEL classification: D62, F53, H23, Q55

1 Introduction

Countries have been struggling for over twenty years to reach an effective international agreement to reduce emissions of greenhouse gases. Even if current negotiations are successful, and more countries commit to meaningful emission reductions than under the Kyoto Protocol, the new agreement will not enter into force before 2020. In the meantime, countries have to make important decisions about their technology policy.

In this paper, we analyze the interaction of abatement and R&D efforts in the area of climate cooperation. We treat climate stability and knowledge as two global public goods, and assume that each country can contribute to both of them. In line with Barrett (1994) and other authors from the climate cooperation literature, we assume that countries decide individually and non-cooperatively whether they want to join a coalition or not.[1] Furthermore, R&D investments are determined *before* countries play the climate coalition formation game. Intuitively, the development of low-carbon technologies is a time-consuming process. Therefore, strategic effects upon later abatement arise when countries determine their R&D investments.

Buchholz and Konrad (1994), Beccherle and Tirole (2011), and Harstad (2012) have shown that these strategic effects can negatively affect welfare, because countries tend to reduce their R&D efforts in order to be assigned a lower abatement target at the cooperation stage. In our model, this effect does not arise because R&D affects all countries' abatement cost functions equally due to its public good nature. However, the positive effect of R&D investments on abatement targets leads to a different problem. It makes *participation* in a future climate coalition less attractive. This effect arises even if countries can determine their R&D efforts cooperatively, as long as there is a delay.[2]

An important feature of our model is that it allows for border carbon adjustments (BCA). This reflects the importance that this topic has received in recent discussions of climate agreements (see, e.g., Böhringer, Balistreri and Rutherford, 2012). BCA is a policy instrument designed to prevent competitive disadvantages for firms located in countries that implement unilateral climate policies. Based on a simple trade model, we assume that firms in signatory countries that export their output to non-signatories are partially exempted from the emissions tax in their home country. Furthermore, import tariffs adjust the total emissions price of firms located in non-signatory countries to the tax raised by signatories.

The introduction of BCA substantially changes the outcome of the climate cooperation

[1] Our basic cooperation model builds on a framework developed by Hoel (1992), Carraro and Siniscalco (1993), and Barrett (1994). This framework has subsequently been extended and modified by a number of authors (e.g. Hoel and de Zeeuw, 2010). It combines elements of cooperative and non-cooperative game theory.

[2] Past climate negotiations focused primarily on the issue of emission reductions, and to a lesser extent on the development of low-carbon technologies. A possible explanation may be that the latter are more difficult to monitor than emission levels.

game. Intuitively, BCA reduces the abatement cost of signatories (due to export exemptions) and raises the costs of non-signatories (due to import tariffs). As a consequence, BCA allows signatories to shift some of their abatement activities to non-signatories. In addition, signatories receive transfers from non-signatories via BCA tariffs. All these effects make the option to become a signatory more attractive, which yields a stabilizing effect upon cooperation. The endogenous coalition size can, thus, be higher than in a standard cooperation model without BCA. This allows us to derive a non-trivial relation between countries' R&D efforts and the endogenous coalition size.

In our analysis we neglect the usual integer constraint on the number of signatories. Instead, we treat the participation level in the coalition as a continuous variable. It turns out that this substantially facilitates the determination of the coalition size. This allows us to use a non-parametric approach, and to carry out most of the analysis without closed form solutions and numerical simulations, on which much of the existing literature has relied. Moreover, we will show that this approach is conservative in the sense that it (weakly) underestimates equilibrium participation as obtained in the 'standard' coalition model. Indeed, without BCA our continuous approach would always lead to zero participation. Intuitively, all countries benefit equally from abatement, but costs of abatement are larger for coalition members. Hence, they would always prefer the position of a non-signatory. Accordingly, it is the discontinuity that arises from the integer effect which leads to the usual result of a non-empty coalition. In contrast, in our continuous participation model it is BCA that allows for a positive coalition size.

This enables us to analyze the following questions. Assuming that climate cooperation is currently not feasible – how does the prospect of future cooperation affect countries' current incentives to invest in R&D? And conversely: how do countries' prior investments in R&D affect the endogenous participation in a future climate agreement and the amount of emissions reductions achieved?[3]

If the coalition size were fixed, higher R&D investments would lead to a higher abatement target implemented by a coalition. However, this makes it less attractive to become a coalition member in the first place. As a result, participation decreases, which in turn has a negative effect on the coalition's abatement target. We show that early R&D efforts are often detrimental to welfare – even if R&D costs are neglected. Countries, therefore, have an incentive to reduce their R&D efforts as long as they anticipate that not all countries will participate in a future climate coalition. This underinvestment persists even if countries can cooperate in the R&D dimension (despite the public good nature of R&D in our model). The outcome is thus often characterized by *full participation* in the climate coalition, but the welfare improvements that this grand coalition achieves are limited due to technology underinvestments.

[3]In equilibrium, of course, these two questions cannot be answered in isolation, as countries are forward-looking at the R&D stage, and anticipate participation and abatement by the future coalition.

In an extension of our model (see Section 4), we further show that an alternative outcome may exist, where the free-rider effect manifests itself in the membership game. In this case, countries invest more in R&D, but the welfare gains are now limited due to low participation (as in the classical model by Barrett, 1994). When both types of outcomes coexist, equilibrium choice amounts to a coordination problem at the R&D stage.

Related literature

A significant part of the game-theoretic literature on international environmental agreements focuses on pure abatement games, without incorporating R&D efforts explicitly into the analysis.[4] Jaffe, Newell and Stavins (2005) highlight the relevance of a double public goods problem of emissions reduction and technology provision in the area of climate policy. Recently, progress has been made to improve our understanding of the multiple externality problem of climate change.[5]

Barrett (2006), e.g., analyzes technology provision in a framework that addresses this double public goods problem by a system of two treaties: a technology and an adoption treaty. He considers 'breakthrough technologies' and shows that the free-rider incentive undermines the scope of cooperation, unless the new technologies exhibit 'increasing returns to adoption' (which the author argues is unlikely to be the case for the most relevant low-carbon technologies).[6] Similar to that paper, our model allows for the possibility of an early R&D coalition (although the non-cooperative R&D case that we also consider is probably more relevant). However, an emission abatement coalition forms only at a later stage, taking countries' prior R&D investments as given. In a simultaneous move game without cooperation, Heal and Tarui (2010) analyze how the degree of technology spillovers affects countries' investment decisions in low-carbon technologies.[7]

Various modifications of climate cooperation models have been introduced in the literature. For instance, Barrett (1997) assumes that countries can link climate agreements to trade issues, which allows the signatories of a climate treaty to impose punishments on non-signatory countries. These punishments make the option to join the treaty rel-

[4]For an overview, see Barrett (2005), and Finus (2008).

[5]Goulder and Mathai (2000) analyze the impact of induced technological change on the optimal time path of the emissions price and abatement in a single-country framework. The authors distinguish between technological change based on R&D efforts, and technological change based on learning-by-doing as a byproduct of abatement activities. See also Gerlagh, Kverndokk, and Rosendahl (2009).

[6]Hoel and de Zeeuw (2010) show that more optimistic results can be obtained when the fixed development cost in Barrett (2006) is replaced by a continuous adoption cost function that declines in the total R&D effort.

[7]They find that R&D investments can be higher than under first-best when marginal abatement costs are increasing in R&D efforts and technology spillovers are small. In our model, we focus on the case of knowledge as a pure public good (large technology spillovers), and assume that marginal abatement costs are declining in the aggregate R&D effort. The counter-intuitive result of Heal and Tarui (2010) can, thus, not arise.

atively more attractive, which increases the scope for cooperation.[8] While punishments via trade sanctions may not be in conformity with WTO rules, the BCA measures that we propose are a softer instrument with the goal to level the playing field for international trade when some countries are more active in the area of climate protection than others. Hence, they may be easier to justify than trade sanctions. Moreover, Böhringer, Fischer and Rosendahl (2014) compare different policy instruments and find that BCA ranks first in cost-effectiveness. They also analyze the interaction between BCA and coalition sizes, but the latter is not determined endogenously as in our paper.

Harstad (2012) uses a model with infinitely many periods to study binding climate agreements under dynamic interaction. Instead of a single climate agreement, countries negotiate agreements repeatedly. Similar to our model, countries can invest in R&D, and then cooperate in their abatement efforts. To obtain tractable results, Harstad (2012) assumes an additive structure, in which R&D has no effect upon the marginal costs of abatement. Beccherle and Tirole (2011) allow for a more general relation between R&D investments (or other types of early efforts) and abatement costs, but do not account for repeated interaction. Our model is closely related to this paper. However, while Beccherle and Tirole (2011) focus on the effects of bargaining and abstract from the free-rider incentive, we endogenize participation in the climate coalition. Battaglini and Harstad (2012) also analyze participation in climate agreements, and allow for an endogenous length of the commitment period in a framework with repeated interaction. The authors find that countries' inability to specify R&D efforts in low-carbon technologies in a climate treaty may raise the efficiency of the final outcome.

The remainder of this paper is organized as follows. In Section 2, the cooperation model with BCA is introduced, and equilibrium participation and abatement are analyzed. Section 3 endogenizes countries' non-cooperative R&D efforts, and shows that countries often underinvest in R&D in order to increase participation. In Section 4, a modified version of the model is analyzed. This serves as a robustness check, and yields additional insights. Proofs are relegated to an Appendix.

2 Coalition model with BCA

Consider N symmetric countries, indexed $i = 1, ..., N$, that play the following climate cooperation game: First, countries non-cooperatively decide whether to join a climate treaty. Second, signatories collectively decide about their climate policies so as to maximize their joint welfare. Third, non-signatories individually decide about their climate policies. We now extend this standard model by introducing BCA.

[8] Hoel and Schneider (1997) assume that countries that stay out of a climate coalition incur a social cost of non-cooperation, which also tends to stabilize cooperation.

The main justification for BCA is to balance distortions in countries' international competitiveness. These arise, if some of them unilaterally contribute more to the global public good of climate stabilization than others. We focus on 'full BCA' which results in a destination-based carbon pricing (e.g., Böhringer et al., 2012). Specifically, emissions embodied in exports to a country with a lower carbon price are taxed by the exporting country only at the price of the foreign country. Moreover, emissions embodied in imports from a country with a lower carbon price are subject to a tariff such that the overall carbon tax of the foreign firm matches that of the importing country.

The partial export exemptions reduce abatement of coalition members, while the BCA-tariff raises abatement of non-signatories. Moreover, we assume that BCA tariff revenues accrue to the country that levies the tariff, and thus add to the welfare of signatories.[9] Intuitively, all three effects tend to stabilize cooperation. As a result, the range of coalition sizes in our model is substantially larger than in the standard model, where participation does not exceed three countries for typical specifications of benefit and abatement cost functions (Barrett, 1994). This is a basic prerequisite for analyzing the effects of R&D investments on the coalition size in Section 3.

Since countries are ex-ante identical, it is natural to assume that their export shares are the same, and that exports go in equal proportions to all other countries. In order to keep the analysis tractable, we assume that these export shares of the "polluting goods" are not affected by climate policies *provided that BCA is implemented*. This assumption can be interpreted as the most parsimonious way to capture the idea that BCA achieves its goal of leveling the playing field for trade between signatories and non-signatories. It may be seen as a benchmark for a world in which trade flows are not strongly affected by unilateral climate policies when they are cushioned by BCA. Calibrated numerical simulations show that BCA may indeed constitute an effective instrument for maintaining competitiveness of energy-intensive and trade-imposed (EITE) industries. In a model comparison of the Energy Modeling Forum, the loss of EITE production in abating countries falls on average from 2.8% to roughly 1% (Böhringer et al., 2012, p. 102).

Constant export shares can also be deduced from a highly stylized trade model, where such an outcome arises endogenously. We now present such a trade model. In the context of this model, we will also derive countries' abatement cost functions under BCA. However, we wish to emphasize that the subsequent analysis depends on the assumption of constant export shares, but *not* on the specific microfoundation that we now present and that has been chosen mainly for its simpleness. In more elaborate models, additional effects – e.g., related to the terms-of-trade – would arise. Analyzing such effects would require a substantially more detailed description of the economy, which would render the

[9] This is the default scenario that has been used in the model comparison of the Energy Modeling Forum (Böhringer et al., 2012).

analysis intractable, especially when we later endogenize R&D efforts.[10]

We assume that in each country there is a representative consumer who desires to consume a certain bundle of goods. Furthermore, in each country, there is a unit mass of polluting firms, some of which export their goods to other countries. For parsimony, suppose that each polluting firm produces either for local consumption or for export, but not for both. Furthermore, let demand of each representative consumer for polluting goods be price-inelastic. Specifically, we assume a unit demand with a maximum willingness-to-pay that is larger than production costs plus any emission costs that may arise from carbon pricing.

Hence, consumption patterns of the polluting goods are fixed. Furthermore, with BCA the carbon price that a firm faces depends only on where the good is consumed, but not on where it is produced. This prevents that firms want to relocate in response to different carbon prices across countries.[11] Given that consumption and production patterns of the polluting goods are unaffected by carbon prices, it follows immediately that also trade flows of these goods stay the same. We normalize the total output of polluting goods of each country to 1, and denote total exports of a country to *each* of the other countries by μ. Hence, a country's total output of polluting goods that is exported to the other countries is $(N-1)\mu$, and the remaining output, $1-(N-1)\mu$, is consumed domestically. Note that, by construction, $\mu \in [0, 1/(N-1)]$. We slightly restrict this range by assuming $\mu < 1/N$, which says that exports to each *individual* country do not exceed production for the domestic market.

Now suppose that each firm has access to abatement technologies, and its 'abatement potential' is proportional to its output. More formally, let a_{il} be the *abatement intensity* of a firm l in country i, i.e., its abatement of emissions per unit of output. We assume that a firm's total abatement costs are simply the product of the firm's output and a function $c(a_{il})h$ of the firm's abatement intensity, where h is a scaling parameter.[12] We assume that $c(\cdot)$ satisfies the standard assumptions $c'(\cdot) > 0$, $c''(\cdot) > 0$ for all $a_{il} > 0$, as well as $c(0) = c'(0) = 0$ and $c''(0) \geq 0$.

Countries regulate their emissions via carbon prices, where $p_i \geq 0$ is the price implemented by country i. Accordingly, with BCA all firms that sell their goods in country i face this price and, therefore, choose an identical abatement intensity according to $p_i = c'(a_{il})h$. Moreover, under BCA the total mass of firms affected by p_i is 1. This holds, even though some firms in country i export their good to a foreign country $j \neq i$, and face an emissions price of p_j accordingly. The reason is that, by symmetry of the

[10]Recently, Eichner and Pethig (2013) integrated international trade in a standard coalition model. However, they rely on numerical simulations to solve their model, and do not account for endogenous R&D efforts and BCA.

[11]Firm relocation or more generally the relocation of productive activities is an important channel of 'carbon leakage'. See, e.g., Babiker (2005).

[12]Parameter h will be specified further in Section 3 where countries' R&D efforts are endogenized.

model, there exists an identical number of firms outside of country i that are taxed according to p_i because their output is imported by country i. In conclusion, total abatement induced by country i's carbon price is $a_i \equiv \int_0^1 a_{il} dl = a_{il}$ so that

$$p_i = c'(a_i)h \tag{1}$$

implicitly determines a_i as a function of p_i. In the following, we refer to a_i as country i's *abatement target*. This reflects that the firms located in country i undertake only a fraction $1-(N-1)\mu$ of the actual abatement induced by country i's carbon price. These are the firms that are located in country i and produce for local consumption. Conversely, the exporting firms are subject to the carbon price that prevails in their export markets. This yields the following expression for country i's welfare:

$$\Pi_i = bA - [1-(N-1)\mu]c(a_i)h - \sum_{j \neq i} \mu c(a_j)h + t_i. \tag{2}$$

The first term on the right-hand side denotes the benefits of climate polices, which are assumed to be linear in aggregate abatement, denoted $A \equiv \sum_i a_i$.[13] The next two terms represent abatement costs of firms in country i that sell their output to local consumers (second term) and to foreign countries (third term). Specifically, firms in country i that export to a country j are affected by the foreign carbon price p_j and abate according to $p_j = c'(a_j)h$. Since there is a fraction μ of them in country i, their overall abatement costs are $\mu c(a_j)h$. Finally, net transfers of BCA tariffs to country i are denoted by t_i (see below).

For later reference, it is useful to specify the welfare function for the situation of k signatories (subscript s) and $N-k$ non-signatories (subscript n) that choose identical carbon prices of p_s and p_n, respectively. In this case, welfare of a signatory equals

$$\Pi_s = bA - [1-(N-k)\mu]c(a_s)h - (N-k)\mu c(a_n)h + t_s. \tag{3}$$

The third term on the right-hand side reflects that exports to the $N-k$ non-signatories are subject to their carbon price p_n and, therefore, associated abatement follows from $p_n = c'(a_n)h$. The remaining output, $1-(N-k)\mu$, is subject to the carbon price of signatories, p_s, and associated abatement follows from $p_s = c'(a_s)h$.

Similarly, welfare of a non-signatory is

$$\Pi_n = bA - (1-k\mu)c(a_n)h - k\mu c(a_s)h + t_n. \tag{4}$$

[13]This simplification is widely used in the literature (e.g., Barrett, 2006). Finus et al. (2006) show that discounted climate damages that are linear in emissions are indeed a good approximation of the figures in the RICE model (Nordhaus and Yang, 1996), although damages are non-linear in temperature change.

It remains to determine BCA-induced transfers. Denote baseline emissions per country (and, hence, per polluting firm) by $\bar{E}$, and assume that they are sufficiently large such that they *always* exceed equilibrium abatement a_i. Each non-signatory has to pay the carbon tariff, $p_s - p_n$, for its emissions embodied in exports to the k signatories, $k\mu(\bar{E} - a_s)$. It follows that

$$t_n = -k\mu(\bar{E} - a_s)(p_s - p_n). \tag{5}$$

Such transfers are paid by $N - k$ non-signatories, and shared among k signatories, yielding

$$t_s = (N - k)\mu(\bar{E} - a_s)(p_s - p_n). \tag{6}$$

Finally, observe that for $\mu = 0$ – i.e., if there is no trade and, therefore, no BCA – the welfare function has the 'standard' form $\Pi_i = bA - c(a_i)h$.

2.1 Choice of abatement

We now analyze countries' emission reduction policies. From (1), these can be described equivalently as choosing a carbon price p_i, or an abatement target a_i. In order to keep the model tractable, we make the following assumption.

Assumption 1. Non-signatories do not implement emission reduction policies, i.e. $p_n = 0$.

Accordingly, non-signatories do not only ignore the externality of their emissions upon other countries, but they also ignore negative effects for themselves. This assumption has also been employed in other papers on BCA (e.g., Böhringer et al., 2014). We will relax it in Section 4, where we parameterize the abatement cost function so as to obtain closed form solutions.

Notwithstanding that Assumption 1 has mainly been introduced to keep the general model tractable, it is not so unrealistic after all. In a setting with identical countries, the effect of any individual country on climate damages is small. Moreover, establishing an emissions control scheme is likely to be accompanied by some fixed costs, and potentially faces opposition from lobbying groups (see Sterner and Isaksson, 2006). Therefore, governments may well decide against establishing domestic climate policies.[14]

The assumption $p_n = 0$ implies that $a_n = 0$ (from (1) and $c'(0) = 0$). Thus, to simplify notation, in the following we write p instead of p_s, and a instead of a_s. The

[14]In the standard coalition model without BCA and R&D investments, $p_n = 0$ would actually arise endogenously if one defined baseline emissions $\bar{E}$ as the Nash equilibrium in emissions. Given linear damage cost, abatement of coalition members would have no effect on emissions of non-signatories. For another coalition model where non-signatories always abate zero see Karp and Simon (2013, p. 330). Obviously, the same outcome occurs in models with binary abatement choices, i.e., where countries either abate or pollute (e.g., Barrett, 2001).

welfare function of a signatory (3) and a non-signatory (4) then becomes (using (5) and (6))

$$\Pi_s(k,a) = bA - (1-\mu(N-k))c(a)h + \mu(N-k)p(\bar{E}-a), \text{ and} \tag{7}$$
$$\Pi_n(k,a) = bA - k\mu c(a)h - k\mu p(\bar{E}-a). \tag{8}$$

Turning to emission reduction policies of signatories, we adopt the standard assumption that they choose the carbon price p so as to maximize their joint welfare. However, in our model welfare includes BCA-induced transfers from non-signatories to signatories. This leads to an additional strategic effect as signatories could raise tax revenues from firms outside the coalition by setting a higher carbon price. Such a strategic manipulation of environmental policies in order to raise income from BCA would probably violate WTO rules. This is reinforced by the fact that the coalition acts like a cartel.[15] Therefore, we make the following assumption.

Assumption 2. When coalition members choose their emission reduction policy, they neglect effects on BCA-induced transfers.

Given this assumption, signatories ignore the last term in (7) when choosing their abatement target, a. Hence, maximizing welfare of a coalition of size $k > 0$ yields the first-order condition (the second-order condition is trivially satisfied)

$$bk = [1-\mu(N-k)]c'(a^*)h. \tag{9}$$

This defines the equilibrium abatement target, a^*, as a function of k, yielding $a^*(k)$ and $A^*(k) = ka^*(k)$. We thus obtain the well-known property that coalition members choose abatement such that aggregate marginal benefits of signatories equal their individual marginal costs. However, $1-\mu(N-k) < 1$, so that the carbon price, $p^* = c'(a^*)h$, is larger than in a situation without BCA ($\mu = 0$). This reflects that BCA shifts part of the abatement costs to non-signatories, which gives signatories an incentive to implement a larger carbon price.

Raising the coalition size reduces this effect, ceteris paribus inducing less abatement. However, as k rises more damages are internalized, inducing more abatement. The following result shows that the latter effect dominates.

Lemma 1. *The abatement target $a^*(k)$, and hence total abatement, $A^*(k) = ka^*(k)$, are strictly increasing in participation, k.*

[15] The fear that BCA could be used for back-door trade policy has indeed been emphasized in the literature, and is one of the main political obstacles for its implementation (see Bhagwati and Mavroidis (2007) for a discussion of legal and policy aspects).

2.2 Membership game

We now turn to the first stage of the game, where each country decides whether to become a coalition member. Note that this decision will take into account how BCA-transfers affect welfare. The usual approach is that the equilibrium participation level k is constrained to be an integer that satisfies the following conditions of 'external' and 'internal' stability (e.g., Barrett, 2005):

$$\Pi_n(k, a^*(k)) \geq \Pi_s(k+1, a^*(k+1)), \text{ and } \Pi_s(k, a^*(k)) \geq \Pi_n(k-1, a^*(k-1)). \quad (10)$$

The integer constraint on k makes a general analysis of this game tedious, explaining the widespread reliance on parametric examples in the literature. In order to overcome this problem, we treat the participation level k as a continuous variable. This has also been done in Karp and Simon (2013), but in contrast to them we will introduce a simpler definition of coalitional stability, which corresponds to the case of a continuum of countries, but delivers a good approximation of the coalition size with discrete countries.[16] Below, we show that this approach is conservative in the sense that it tends to underestimate equilibrium participation.

With participation k as a continuous variable, it becomes less obvious for which values of k one can reasonably speak of a 'climate coalition'. In line with the standard approach, this would be the case for $k \geq 2$. However, for the remaining analysis it is more convenient to adopt a broader perspective and refer to all $k > 0$ as a coalition.[17] Accordingly, three different types of outcomes can occur at the participation stage. The two corner solutions $k^* = 0$ and $k^* = N$, as well as *interior solutions*, $0 < k^* < N$.

Definition 1. *Consider continuous participation decisions, $k \in \mathbb{R}_+$. A coalition with $k \in (0, N)$ members is a stable equilibrium iff*

$$\Pi_s(k, a^*(k)) = \Pi_n(k, a^*(k)), \text{ and} \quad (11)$$

$$\tfrac{d}{dk}\left[\Pi_s(k, a^*(k)) - \Pi_n(k, a^*(k))\right] < 0. \quad (12)$$

The grand coalition, $k = N$, is a stable equilibrium iff either (11) *and* (12) *hold, or*

$$\Pi_s(N, a^*(N)) > \Pi_n(N, a^*(N)). \quad (13)$$

Conditions (11) and (12) are the continuous equivalent of the standard stability con-

[16] Martimort and Sand-Zantman (2013) assume a continuum of countries in their analysis of climate agreements. Haufler and Wooton (2010) analyze a location game, where (similar to our approach) the number of *firms* that decide to locate in each country is treated a continuous variable.

[17] One could think of coalitions $0 < k < 1$ as situations where only some regions in a country implement climate policies, e.g., California in the US. However, note that this choice is one of exposition and not of substance. All results remain valid if one restricts coalitions to $k \geq 2$.

ditions (10).[18] They assure that (*i*) a non-signatory does not want to become a signatory as the associated increase in k would lead to $\Pi_s < \Pi_n$, and (*ii*) no signatory wants to leave the coalition as it would then receive $\Pi_n < \Pi_s$.[19] Condition (13) reflects that the grand coalition is also stable if the payoff of a coalition member is strictly larger than the payoff which it would obtain from leaving the coalition.

Proposition 1. *Consider a participation level $k \in (0, N]$ that satisfies* (11)*. If $\mu N a + c'(a)/c''(a)$ is (weakly) increasing in a, then k is the unique interior solution of the continuous membership game, i.e., it also satisfies* (12)*.*

This reflects the well-known result that the incentives to become a free-rider are increasing in the coalition size. Therefore, when the welfare of signatories and non-signatories is equal at k^*, then for any $k > k^*$ non-signatories are strictly better off. Hence, signatories would prefer to leave the coalition.

The condition in Proposition 1 is equivalent to

$$1 + \mu N - \frac{c'(a)c'''(a)}{(c''(a))^2} \geq 0.$$

It is satisfied when $c'''(a) \leq 0$, or when $c'''(a) > 0$ but the third-order effect is not too strong. Moreover, for an isoelastic cost function

$$c(a) = a^z/z, \ z > 1, \tag{14}$$

the condition is always fulfilled since $c'(a)/c''(a)$ is increasing in a.[20] Note, that the isoelastic case includes the widely used quadratic specification (obtained for $z = 2$). Hence, the condition does not appear to be overly restrictive.

In the remaining, we assume that it is satisfied. As a consequence, determination of interior participation levels is straightforward for the continuous approach. One only has to solve condition (11) for k. The next proposition shows that the resulting participation level provides a conservative estimate of the coalition size that would obtain if the

[18] Karp and Simon's (2013) approach is to determine k such that $\Pi_s(k, a^*(k)) = \Pi_n(k-1, a^*(k-1))$. Even when k is treated as a continuous variable, this takes into consideration that a deviation by a country at the participation stage leads to a non-marginal change in the size and, therefore, abatement of a coalition. This makes it more costly to leave a coalition. Our approach ignores this effect, which explains why it underestimates the coalition size.

[19] If there exists a unique solution $k^* \in (0, N)$ of (11) that violates (12), and $\Pi_s(k, a^*(k))$ and $\Pi_n(k, a^*(k))$ are increasing in k for all $k > 0$, then k^* is a separation point between two cases: for $k < k^*$ it holds that $\Pi_s < \Pi_n$, so signatories have an incentive to leave the coalition; for $k > k^*$ it holds that $\Pi_s > \Pi_n$, so non-signatories prefer to join the coalition. Consequently, there are two equilibria at the participation stage: $k = 0$ (no coalition forms) and $k = N$ (grand coalition). In this case, countries face a coordination problem. Only if a sufficient number of countries joins the coalition, participation becomes profitable for all countries.

[20] In particular, $c'(a)/c''(a) = (z-1)a$. The parameter z is the elasticity of the cost function, and $z > 1$ assures that $c(\cdot)$ is convex.

standard integer constraint were added to the problem.

Proposition 2. *Denote by $I(k)$ the largest integer smaller than or equal to k. Let k^* be the unique solution of the membership game with continuous participation decisions. Any integer k' that solves the internal and external stability conditions* (10) *is at least as large as $I(k^*)$.*

Without BCA the payoff of a non-signatory would always be larger than the payoff of a coalition member because non-signatories share the same benefits of abatement but incur lower costs. Hence, condition (11) could not be satisfied for any $k > 0$, and the unique equilibrium outcome with continuous participation would be $A^* = k^* = 0$. BCA changes this picture significantly, and allows for higher participation in our model.

In particular, from Definition 1 and Proposition 1, the grand coalition is stable if $\Pi_s(N, a^*(N)) \geq \Pi_n(N, a^*(N))$. Substituting from (7) and (8), thereby using (1), the condition becomes

$$(N\mu - 1)c(a^*(N)) + N\mu c'(a^*(N))(\bar{E} - a^*(N)) \geq 0. \quad (15)$$

Note, that $a^*(N)$ is independent of μ, because BCA plays no role when all countries cooperate. The left-hand side is, thus, increasing in μ. Moreover, the first (negative) term converges towards 0 as μ converges to its upper limit, $1/N$. Since the second term is strictly positive, we obtain the following result.

Proposition 3. *There always exists an export share $\mu < 1/N$ such that the grand coalition obtains with BCA.*

Intuitively, the more output of polluting goods is exported to signatories – and, therefore, subject to BCA – the smaller the advantage from being a non-signatory. Moreover, observe that whenever the grand coalition is an equilibrium in the model with continuous participation decisions, then it is also an equilibrium according to the "standard" stability condition (10) (this follows immediately from $I(N) = N$ and Proposition 2).

The solid line in Figure 1 depicts k^* (for continuous participation decisions) as a function of μ for the following specification: $h = 1, N = 100, b = 1, c(a) = 0.5a^2$, and $\bar{E} = 140$.[21] It shows (for this specification) that already low export shares are sufficient to trigger coalitions that are substantially larger than in the standard model. The grand coalition obtains once μ has reached a level that is slightly larger than half of its maximum value, $1/N = 0.01$.

The step function in Figure 1 depicts the equilibrium participation level for the standard integer approach, i.e. based on conditions (10) (rather than (11)). The figure does

[21] Equilibrium abatement in the grand coalition is $a^*(N) = 100$. Hence $\bar{E}$ has been chosen such that emissions are reduced by roughly 70% in the grand coalition.

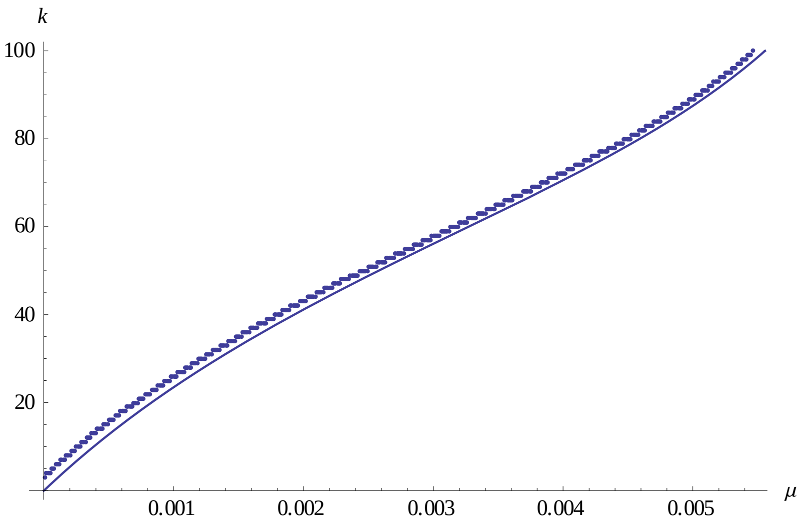

Figure 1: Effects of BCA on coalition size

not only confirm the continuous approach as being more conservative (Proposition (2)), but also shows that both approaches do actually lead to quite similar results.[22]

3 Climate cooperation with endogenous R&D

The climate cooperation game with BCA that was introduced in the previous section allows for a non-trivial amount of participation in equilibrium. We now extend this model and assume that countries can invest in low-carbon technologies in order to reduce their abatement costs. Following Buchholz and Konrad (1994) and Beccherle and Tirole (2011) we assume that there is an *exogenous delay* in countries' efforts to cooperate on emission mitigation. To formalize this idea, we add a new stage to our climate cooperation model. At this prior R&D stage ('stage 0'), countries determine their R&D efforts. For most parts of this section, it does not matter whether countries do so cooperatively or non-cooperatively because the results depend only on the aggregate R&D level.

Let $r_i \geq 0$ be country i's R&D effort. It is associated with a cost $f(r_i)$ that satisfies $f(r_i) > 0$, $f'(r_i) > 0$, $f''(r_i) > 0$ for all $r_i > 0$, as well as $f(0) = f'(0) = 0$, and $f''(0) \geq 0$. R&D efforts reduce the abatement cost parameter h. We assume that R&D is – like abatement – a pure public good so that h is now a *function* of the aggregate R&D level, denoted $R \equiv \sum_i^N r_i$. This reflects that technological knowledge can only partially be appropriated by those who invest in it, due to knowledge spillovers. If

[22]Note, that without BCA, i.e. for $\mu = 0$, our model predicts $k^* = 3$ as the unique solution for the integer approach. By contrast, in the standard model (e.g., Barrett, 1994) also $k^* = 2$ is an equilibrium. This difference is due to our assumption that non-signatories set $p_n = 0$. Intuitively, if non-signatories are restricted to abate less than would be optimal for them, it becomes more attractive to join the coalition.

these are significant, as suggested by a number of studies (see e.g. Coe et al., 2009, Lee, 2003), then most of the knowledge developed by a single country becomes available to all countries. The simplifying assumption that R&D is a *pure* public good highlights the problem of knowledge spillovers, and may be a reasonable approximation especially when the time span under consideration is long.[23] Moreover, we assume a multiplicative specification of abatement costs, $c(a_i)h(R)$, where $h(R) > 0$, $h'(R) < 0$ and $h''(R) > 0$ for all $R \geq 0$, and $\lim_{R\to\infty} h(R) = 0$. This has some nice properties which we exploit in the following analysis.

3.1 Effects of R&D investments on coalition game

For a given coalition size k, the identity of the countries that enter the coalition is randomly determined. Therefore, at the R&D stage the expected payoff per country is $\pi(R) \equiv \frac{k}{N}\Pi_s + \frac{N-k}{N}\Pi_n - f(r_i)$. Moreover, since countries are ex-ante identical, they will choose identical R&D efforts so that $r_i = R/N$. Using (7) and (8) this yields

$$\pi(R) = bA^*(R) - \frac{k^*(R)}{N}c(a^*(R))h(R) - f(R/N), \tag{16}$$

where $A^*(R) \equiv A^*(R, k^*(R))$ is total abatement (as function of R) when participation $k^*(R)$ and the abatement target $a^*(R) \equiv a^*(R, k^*(R))$ are determined as the equilibrium outcome of the cooperation game (stages 1 and 2) given R.[24] Note, that transfers from non-signatories to signatories dropped out, as ex-ante *expected* transfers are zero for each country. We obtain the following results for variations in the aggregate R&D effort, R.

Proposition 4. *For interior solutions at the participation stage, the abatement target $a^*(R)$ is independent of the R&D level R, and the coalition size $k^*(R)$ is decreasing in R. Therefore, higher R&D investments lead to less overall abatement, $A^*(R)$. Furthermore, let ε denote the elasticity of the cost function $c(\cdot)$. If it satisfies*

$$\varepsilon \geq \frac{k}{N[1-\mu(N-k)]}\left(2 - \frac{\mu k}{1-\mu N+\mu k}\right), \tag{17}$$

then each country's expected payoff, $\pi(R)$, is decreasing in R.

[23] The pure public good assumption assures that all countries are identical when they enter the coalition game as analyzed in Section 2. This occurs even if they choose different r_i, which would not be the case if R&D also had private good properties. However, with non-identical countries the analysis of the coalition game becomes less tractable. E.g., one needs to check whether changing an individual r_i constitutes a profitable deviation even when all other countries undertake equal R&D investments. But with non-identical abatement cost functions, our assumption that all countries become signatories with an identical probability (see below) seems less compelling. Hence, countries might choose different R&D efforts, depending on whether they are *planning* to become signatories or non-signatories. This complicates the analysis significantly.

[24] We use the same notation as in the previous section, but add a new argument R to the functions $A(\cdot)$ and $a(\cdot)$ in order to capture the dependency of h from R.

Note, that these results were obtained without imposing any further restrictions on the functions $h(\cdot)$ and $f(\cdot)$. Only the last result concerning welfare effects comes with the caveat of a sufficiently large elasticity of the cost function $c(\cdot)$. Specifically, $k/N[1 - \mu(N-k)] \leq 1$ so that the right-hand side of (17) is strictly smaller than 2. Hence a sufficient condition is $\varepsilon \geq 2$, which includes the standard case of a quadratic cost function. Moreover, as k converges towards 0, the right-hand side of (17) also does so. Hence the condition may be satisfied also for substantially lower elasticities. In addition, (17) is itself only a sufficient condition because it neglects R&D costs (see the proof of Proposition 4). Since these are positive, the negative welfare effect of R obtains even for elasticities that violate (17). In conclusion, the caveat appears to be fairly weak. We will assume throughout the rest of this paper that condition (17) is fulfilled.

Intuitively, one would expect that the public good nature of R&D, which reduces the abatement cost of *all* countries, should be beneficial for abatement. Proposition 4 shows that this is not the case if one accounts for the strategic considerations that underlie the formation of a climate coalition. In particular, higher R&D investments reduce marginal abatement costs, which induces signatories to choose a higher abatement target A (given k). Signatories benefit from this, but non-signatories even more because they realize the same benefits of abatement but incur a smaller share of the costs. This makes free-riding more attractive. Therefore, participation declines for higher R&D levels. Due to the smaller coalition size, coalition members choose lower abatement targets (given R). It turns out that for interior solutions of k, these two effects keep each other precisely in balance so that $a^*(R)$ stays constant when R is raised. Given the smaller coalition size k^*, this implies that overall abatement $A = ka$ unambiguously decreases in R.

Turning to the associated welfare effects, there is a trade-off. On the positive side, a higher R&D level reduces abatement costs for any given level of abatement. On the negative side, there are welfare losses due to the lower level of overall abatement, and R&D is costly. The result shows that the negative effects dominate for reasonable specifications of the abatement cost function.

Above we have analyzed for interior participation levels how welfare is affected by changes in the aggregate R&D effort, R. However, if a country's payoff decreases in aggregate R&D, and this is a pure public good (by assumption), then it must also decrease in the individual R&D effort r_i. From the above discussion, this result is driven by the negative effect of R on the coalition size k. Hence, countries have an incentive to lower their R&D efforts because these are detrimental to welfare. Note, that this holds, irrespective of whether the R&D efforts are determined cooperatively or non-cooperatively. However, at a boundary solution $k^* = N$, this effect is absent. Therefore, given our assumption $f'(0) = 0$, countries have an incentive to undertake *some* investments in R&D as long as a boundary solution with $k^* = N$ exists (for sufficiently small values of $R > 0$).

Hence, the next result follows from Proposition 4.

Proposition 5. *There exists no equilibrium that entails both a positive R&D effort R and an interior solution at the participation stage. Hence, an equilibrium with $k^* > 0$ entails either $R^* > 0$ and $k^* = N$, or $R^* = 0$ and $k^* < N$.*

To foster intuition, consider a hypothetical situation with positive R&D levels *and* an interior solution at the participation stage. Reducing R&D investments would raise the costs of achieving a given abatement level. However, it would lower R&D costs, while at the same time raising the coalition size and, thereby, overall abatement (by Proposition 4). The above result shows that the positive effects of reduced R&D efforts dominate until either the grand coalition forms (i.e., $R^* > 0, k^* = N$), or R&D investments cannot be reduced any further (i.e., $R^* = 0, k^* < N$).

3.2 Interior and boundary solutions of R&D

It remains to determine which of the two cases in Proposition 5 applies in equilibrium, and what the associated R&D levels are. Remember that for interior solutions at the participation stage, countries have an incentive to reduce their R&D investments in stage 0 (by Proposition 4), irrespective of whether the R&D efforts are determined cooperatively or non-cooperatively. Suppose there exists an $R \geq 0$ such that the grand coalition obtains in the equilibrium of the climate cooperation game. Denote by R_0 the highest R&D level for which this is the case. It corresponds to the kink in the upper graph of Figure 2.

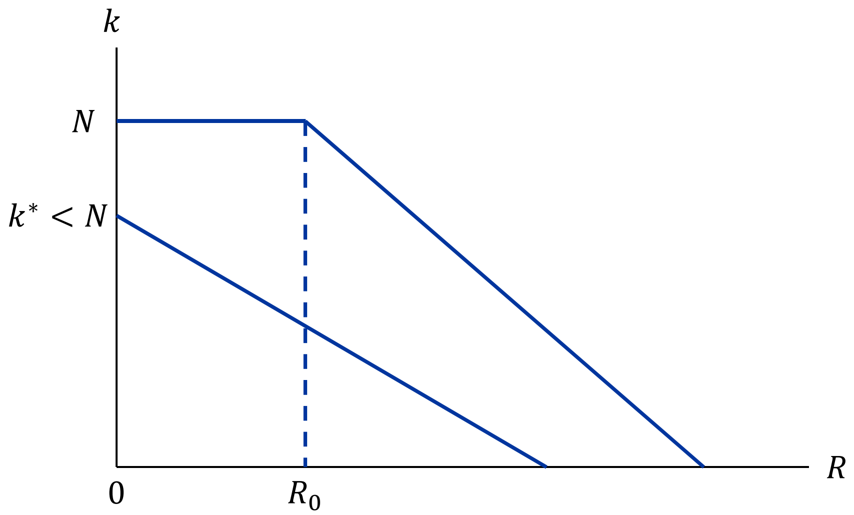

Figure 2: R&D investments and coalition size

From Definition 1 about stability of the grand coalition, it follows that R_0 must satisfy $\Pi_s(N, a^*(R_0), R_0) = \Pi_n(N, a^*(R_0), R_0)$. Moreover, the equilibrium abatement target, $a^*(R_0)$, then follows from evaluating (9) at $k = N$. In sum, R_0 and $a^*(R_0)$ are implicitly determined by the following system (using (7) and (8)):[25]

$$(1 - \mu N)c(a) = c'(a)\mu N(\bar{E} - a), \text{ and} \tag{18}$$

$$bN = c'(a)h(R_0). \tag{19}$$

However, at this level of generality, it is not guaranteed that there exists an $R_0 \geq 0$ that solves the above system. To see this, note that (18) implicitly determines the equilibrium abatement target, independently of R_0. For this value of a, (19) may not have a (non-negative) solution for R_0. Specifically, for $h(0)$ sufficiently small one gets $bN > c'(a)h(0)$, and $h(\cdot)$ is decreasing in R. In this case, the unique equilibrium outcome entails a boundary solution at the R&D stage: $R^* = 0$, and a smaller coalition forms ($k^* < N$). This situation is depicted in the lower graph of Figure 2.[26]

Proposition 6. *The grand coalition forms iff there exists an $R_0 \geq 0$ that solves* (18) *and* (19)*. Otherwise, countries do not invest in R&D ($R^* = 0$), and a smaller coalition forms ($k^* < N$). Furthermore, the critical value R_0 is increasing in μ, and a higher μ makes it more likely that an interior solution $R_0 > 0$ of system* (18) *and* (19) *exists.*

Remember that a higher μ implies that BCA has a greater impact. Hence the last statement in Proposition 6 suggests that BCA has not only a positive effect on the coalition size (Proposition 3), but also on R&D efforts.

Let us conclude this subsection with a characterization of the equilibrium R&D effort R^* when there exists an $R_0 \geq 0$ that solves (18) and (19); i.e., when the grand coalition forms. Clearly, when $R_0 = 0$, then this is also the equilibrium R&D effort. However, when $R_0 > 0$, two cases must be distinguished. To see this, consider the following modified game. Let participation be *exogenously* fixed at $k = N$; i.e., stage 1 of the full game is suppressed. Denote the equilibrium R&D effort in this modified game by R_N. Since strategic effects related to endogenous changes in participation do not arise in this case, R_N generally differs from R_0. If $R_N \geq R_0$, then in the full game the equilibrium R&D effort is R_0, because a higher R&D effort would reduce participation (by definition of R_0) and welfare (by Proposition 4). This implies a 'boundary solution' $R^* = R_0$. However, if $R_N < R_0$, then the equilibrium R&D effort in the full game is equal to R_N, as countries benefit from reducing their R&D effort below R_0. This leads to a solution $R^* < R_0$, which is interior in the sense that R could be raised without compromising the grand coalition.

[25]Note that R_0 is independent of the specification of the function $f(\cdot)$.

[26]Observe that the Inada-type condition $h(0) = \infty$ would assure that (19) has an interior solution for R_0. However, the assumption that marginal abatement costs are infinitely large without R&D investments is not very convincing.

In contrast to R_0, the value of R_N depends on whether countries determine their R&D efforts in stage 0 cooperatively or non-cooperatively. Denote the value of R_N in these two cases by R_N^c and R_N^{nc}, respectively. To determine R_N, note first that the abatement target a follows again from (9). The reason is that when signatories choose their abatement target in stage 2, it is irrelevant whether participation was determined as part of the equilibrium outcome of the full game (in stage 1), or whether it is exogenously fixed (as in the modified game that we are considering here). At $k = N$ this yields the first-order condition

$$bN - c'(a)h(R_N) = 0. \tag{20}$$

Turning to R&D efforts in this modified game, first suppose that these are determined *non-cooperatively*, which is probably the more relevant case. Then for an interior solution, R_N^{nc} follows from maximizing a country's expected payoff as given by (16) at $k^* = N$ with respect to r_i. The resulting first-order condition is (using (20)):[27]

$$-c(a)h'(R_N^{nc}) = f'(R_N^{nc}/N). \tag{21}$$

Alternatively, if R&D efforts are determined *cooperatively*, r_i maximizes the sum of expected payoffs over all countries, yielding the first-order condition for R_N^c (using (20) and noting that the last term in (16) represents $f(r_i)$):

$$-Nc(a)h'(R_N^c) = f'(R_N^c/N). \tag{22}$$

Accordingly, $\{a^*(R_N^c), R_N^c\}$ follow from (20) and (21), and $\{a^*(R_N^{nc}), R_N^{nc}\}$ from (20) and (22). The latter system also characterizes the R&D effort and abatement in the first-best outcome, since both externalities are fully internalized. Obviously, R&D and, therefore, abatement are larger than in the non-cooperative case.

Note that in both cases of cooperative and non-cooperative R&D choices, R_N is independent of $\bar{E}$ and μ, in contrast to R_0. The reason is simply that R_N is the equilibrium R&D level for $k \equiv N$, and when the grand coalition forms there is no BCA so the parameters $\bar{E}$ and μ are irrelevant.

Note further, that the above first-order conditions may not always deliver a (unique) interior solution for R_N^{nc} and R_N^c because the second-order conditions may not be satisfied for the general model specification. A discussion of this problem is in the Appendix (for the non-cooperative case). There, it is also shown that when $c(\cdot)$ is isoelastic with $z \geq 2$ and $h(R) = 1/R$, the second-order condition for R_N^{nc} to be a maximum is always fulfilled.

We summarize our findings in the following proposition.

[27] Note that with $k = N$ fixed, $a^*(R)$ depends on R (in contrast to the result of Proposition 4 which is valid only for an interior solution at the participation stage).

Proposition 7. *If there exists an* $R_0 \geq 0$ *that solves* (18) *and* (19)*, then the equilibrium R&D level is* $R^* = \min\{R_0, R_N\}$*, where* R_N *is the endogenous R&D effort under fixed participation* ($k \equiv N$)*. If R&D efforts are determined non-cooperatively* ($R_N = R_N^{nc}$)*, then free-riding at the R&D stage of the full game only occurs if* $R_N^{nc} < R_0$*. Otherwise (i.e., when* $R_N^{nc} \geq R_0$*), the equilibrium R&D level is* R_0*, independent of whether countries choose R&D cooperatively or non-cooperatively.*

4 R&D with climate policies by non-signatories

The preceding analysis was based on Assumption 1 that non-signatories do not implement own emission reduction policies and set $p_n = 0$. We motivated this assumption by referring to, e.g., fixed costs of introducing an emissions control scheme. However, this is not modelled; to the contrary, we assumed $c(0) = c'(0) = 0$. Given this specification, the assumption $p_n = 0$ imposes a sub-optimal policy upon non-signatories. Relaxing it would raise their welfare and, therefore, make it less attractive to become a signatory. In conclusion, dropping Assumption 1 would lead to smaller coalitions. Nevertheless, qualitatively the results in Section 2 should be preserved as the mechanisms that underlie them do not change. In particular, numerical simulations (using the specification on which Figure 1 is based) show that the curves $k^*(\mu)$ for "$p_n = 0$" and for "p_n endogenous" lie close to each other (available upon request).

Especially for a scenario of low R&D investments this is quite intuitive. In this case emission reductions are expensive and, therefore, non-signatories would only abate little anyway. By contrast, for the analysis in Section 3 with endogenous R&D investments, the implications of Assumption 1 can be more substantial. Given the complexity of the model, general results are hard to obtain. Hence in the following we rely on a simple numerical simulation exercise to illustrate this.[28]

Remember that the results in Section 3 were driven by the effect that higher levels

[28] When non-signatories implement climate policies too, this affects the specification of transfers, depending on the exact way in which BCA is modeled. For the simulation we assume that *all* exports are completely exempted from the local emissions price, and instead taxed in the importing country at the emissions price which prevails there. If non-signatories do not implement climate policies ($p_n = 0$), then this approach is formally identical to our earlier approach because in this case $p_s - p_n = p_s$. However, when $p_n > 0$ the new approach is conceptually simpler because it does not necessitate a case-distinction in transfers, depending on whether the foreign country has a higher or lower emissions price than the home country. The resulting transfers are:

$$t_s = (N - k)\mu \left[p_s(\bar{E} - a_s) - p_n(\bar{E} - a_n)\right], \tag{23}$$

$$t_n = -k\mu \left[p_s(\bar{E} - a_s) - p_n(\bar{E} - a_n)\right], \tag{24}$$

where abatement targets a_s and a_n are implicitly determined by equation (1). Otherwise, the model is solved in the same way as in Section 2. Figure 3 below is based on the following specification: $c(a) = a^2/2, b = \bar{E} = 1, h(R) \equiv 1/R, N = 100$, and $\mu = 0.001$ (so $\mu N = 0.1$, hence, the export share of polluting goods per country is roughly 10%).

of R reduce the coalition size and welfare, even if R&D is costless (Proposition 4). This is represented by the declining, dotted curve in Figure 3, which shows gross welfare (neglecting R&D costs) in the case where $p_n = 0$ (analyzed in Sections 2 and 3). In the boundary solution where no coalition forms, this effect is absent. This explains why the curve becomes less steep for larger values of R.

Now consider the other extreme case, which is the fully non-cooperative benchmark where $k \equiv 0$ is exogenously fixed, but the non-signatories can implement emission reduction policies (case $p_n > 0$). This is represented by the increasing, dotted curve in Figure 3. In this case, costless R&D obviously raises the abatement and welfare of the non-signatories, so that gross welfare per country is strictly increasing in R.

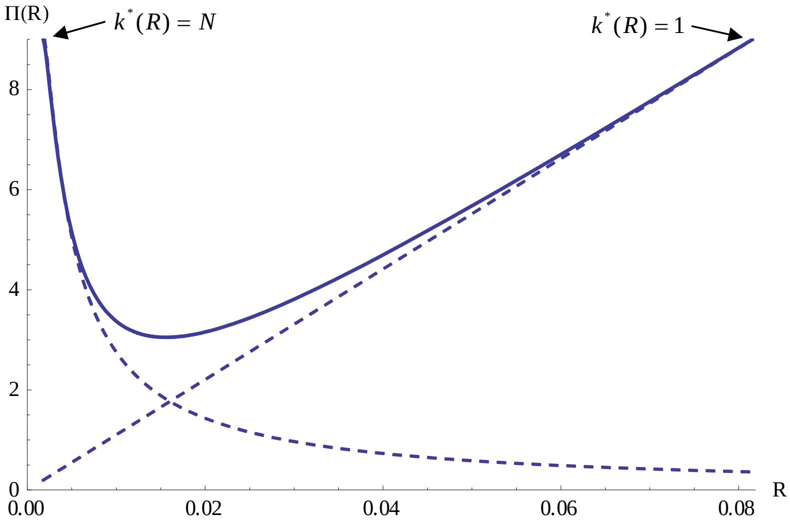

Figure 3: Gross welfare per country as function of R

The U-shaped solid line combines these two countervailing effects and depicts gross welfare per country in the case with endogenous coalitions *and* climate policies implemented by non-signatories (case $p_n > 0$). When R becomes large, participation declines and welfare in the full model converges towards the respective value in the fully non-cooperative benchmark.[29] Conversely, when R becomes small (close to the point where $k^*(R) = N$ holds – hence, close to R_0), results in the full model with $p_n > 0$ almost coincide with those under $p_n = 0$. This is also intuitive, as for R sufficiently small, total abatement of non-signatories becomes negligible so that it makes little difference whether the model allows for climate policies of these countries or not.

Note, that Figure 3 depicts *gross* welfare. When investment costs in R&D are included,

[29]In the case with $p_n > 0$, the range for an interior solution at the participation stage is $k \in (1, N)$, rather than $k \in (0, N)$. This is due to the fact that for $k = 1$, the single signatory behaves in the same way as each of the non-signatories, while for $k < 1$ a signatory would implement a lower emissions price than the non-signatories. The figure shows gross welfare in the range of values of R for which participation lies between $k^*(R) = 1$ (right-hand side) and $k^*(R) = N$ (left-hand side of the figure).

it suggests that two types of outcomes may occur in equilibrium. Either an equilibrium is attained in which R is small ($R = \min\{R_0, R_N\}$), such that $k^*(R) = N$ obtains (full participation). In this case, the main results of Section 3 are preserved when we drop Assumption 1. However, despite the formation of a grand coalition, the realized welfare gains are limited due to significant underinvestment in R&D, which leads to inefficiently high abatement costs. Or, an equilibrium is attained with higher investments in R&D, but low participation. This type of equilibrium can only exist if the investment cost function in R&D, $f(r_i)$, is sufficiently flat so that also *net* welfare would be U-shaped. If both equilibria coexist, then equilibrium selection is a matter of coordination. In conclusion, the free-rider incentive can manifest itself *either* in the R&D-dimension with significant underinvestments in low-carbon technologies but high participation (as in Section 3), *or* in the participation dimension (similarly as in Barrett, 1994), in which case the underinvestment problem is less severe.

5 Concluding remarks

In order to tackle the issue of climate change, countries must undertake costly actions to reduce their greenhouse gas emissions. They must also invest in low-carbon technologies in order to achieve their abatement targets at a reasonable cost. In the light of strong knowledge spillovers, climate change mitigation is thus a global 'double public goods problem'. If a global agreement that fixes countries' abatement *and* R&D efforts is currently not feasible, then the question arises whether early action in the development of low-carbon technologies may pave the way towards a more effective outcome in the future.

Using a stylized cooperation model, we show that border carbon adjustments can be quite effective in addressing the participation problem and guiding countries into a climate coalition. By contrast, our results indicate that early investments in low-carbon technologies may not necessarily improve the outcome of a future climate agreement. In particular, while such investments reduce future abatement costs and trigger additional abatement efforts (for a fixed coalition size), they also aggravate the free-rider incentive, thereby undermining the potential welfare gains of a future climate agreement.

In our model, the crowding-out effect via reduced participation is often so strong that early technology investments are even harmful. Hence, anticipating membership decisions in a future climate coalition as well as its abatement effort, countries strategically underinvest in low-carbon technologies. This way, a future agreement may even reach full participation, but the welfare gains achieved by this grand coalition are limited due to the underinvestment problem in R&D, which implies inefficiently high abatement costs and too little abatement.

This suggest that an effective international environmental agreement on climate change requires high participation, and must fix not only countries' abatement efforts, but also R&D efforts. In our analysis, we did not consider the case where coalitions decide simultaneously about abatement and R&D. However, we conjecture that the free-rider incentive might be even more pronounced. Intuitively, more R&D triggers additional abatement by the coalition, and a higher abatement target makes further R&D efforts by the signatories profitable. Thus the two reinforce each other. By contrast, non-signatories would have little incentive to invest in either of the two public goods, but benefit from the signatories' efforts. This raises the attractiveness of being a coalition outsider, aggravating free-rider incentives. We have seen that BCA may be helpful to address this problem, assuming that it is possible to resolve the political and implementation problems associated with it (which may be substantial). Furthermore, its effectiveness will depend on countries' export shares being large enough.

Despite the negative effects of early efforts to bring down the costs of low-carbon technologies in our model, the results should not be interpreted as saying that these are *generally* counter-productive. Potentially important aspects of the problem such as altruism, a 'true willingness to cooperate' in the sense that some countries may sacrifice more than they can expect to gain from an agreement, or the possibility of a 'technological breakthrough' and increasing returns in the renewable energy sector are omitted. Hence policy implications should be treated with caution. Nevertheless, what our model *does* show is that early investments in technology do not *automatically* lead to a better outcome in future climate negotiations. Climate cooperation remains a challenge, and this challenge may become even bigger if low-carbon technologies are available at lower costs, due to a stronger free-rider incentive. Notwithstanding, one should also keep in mind that the *potential gains* of cooperation are larger when abatement costs are reduced. Hence, our results may challenge policy makers to try even harder to reach an agreement.

References

Babiker, M.H. (2005) Climate change policy, market structure, and carbon leakage. *Journal of International Economics*, 65, p. 421-445

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

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

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

Barrett, S. (2005) The Theory of International Environmental Agreements, in Mäler and Vincent, eds., Handbook of Environmental Economics, 3, Elsevier

Barrett, S. (2006) Climate Treaties and "Breakthrough" Technologies, *AEA Papers and Proceedings*, 96, p. 22-25

Battaglini, M. and B. Harstad (2012) Participation and Duration of Environmental Agreements. NBER Working Paper 18585

Beccherle, J. and J. Tirole (2011) Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements. *Journal of Public Economics* 95, p. 1339-1348

Bhagwati, J. and P.C. Mavroidis (2007) Is action against US exports for failure to sign Kyoto Protocol WTO-legal? *World Trade Review*, 6(2), p. 299-310

Böhringer, C., Balistreri, E.J. and T.F. Rutherford (2012) The role of border carbon adjustment in unilateral climate policy: Overview of an Energy Modeling Forum study (EMF 29). *Energy Economics*, 34(S2), p. S97-S110

Böhringer, C., Fischer, C. and K.E. Rosendahl (2014) Cost-Effective Unilateral Climate Policy Design: Size Matters. *Journal of Environmental Economics and Management*, in press, http://dx.doi.org/10.1016/j.jeem.2013.12.008

Buchholz, W. and K. Konrad (1994) Global Environmental Problems and the Strategic Choice of Technology, *Journal of Economics*, 60, p. 299-321

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

Coe, D.T., Helpman, E. and A.W. Hoffmaister (2009) International R&D spillovers and institutions. *European Economic Review*, 53, p. 723-741

Eichner, T. and R. Pethig (2013) Self-enforcing environmental agreements and international trade. *Journal of Public Economics*, 102, p. 37-50

Finus, M. (2008) Game theoretic research on the design of international environmental agreements: insights, critical remarks, and future challenges, *International Review of Environmental and Resource Economics*, 2, p. 29-67

Finus, M., Ierland. E. and R. Dellink (2006) Stability of Climate Coalitions in a Cartel Formation Game, *Economics of Governance*, 7(3), p. 271-291

Gerlagh, R., Kverndokk, S. and K.E. Rosendahl (2009) Optimal Timing of Climate Change Policy: Interaction Between Carbon Taxes and Innovation Externalities, *Environmental and Resource Economics*, 43, p. 369-390

Goulder, L.H. and K. Mathai (2000) Optimal CO2 Abatement in the Presence of Induced Technological Change. *Journal of Environmental Economics and Management*, 39, p. 1-38

Harstad, B. (2012) The Dynamics of Climate Agreements, working paper, No 2962, CESifo Working Paper Series from CESifo Group Munich, 2010, updated August 2012

Haufler, A. and I. Wooton (2010) Competition for firms in an oligopolistic industry: The impact of economic integration. *Journal of International Economics*, 80, p. 239-248

Heal, G. and N. Tarui (2010) Investment and emission control under technology and pollution externalities, *Resource and Energy Economics*, 32, p. 1-14

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

Hoel, M. and K. Schneider (1997) Incentives to Participate in an International Environmental Agreement, *Environmental and Resource Economics*, 9, p. 153-170

Hoel, M. and A. de Zeeuw (2010) Can a Focus on Breakthrough Technologies Improve the Performance of International Environmental Agreements? *Environmental and Resource Economics*, 47, p. 395-406

Jaffe, A.B., Newell, R.G. and N. Stavins (2005) A tale of two market failures: Technology and environmental policy, *Ecological Economics*, 54, p. 164-174

Karp, L.S. and L. Simon (2013) Participation games and international environmental agreements: A non-parametric model. *Journal of Environmental Economics and Management*, 65, p. 326-344

Lee, G. (2003) The effectiveness of international knowledge spillover channels, *European Economic Review*, 50, p. 2075-2088

Martimort, D. and W. Sand-Zantman (2013) A Mechanism Design Approach to Climate Agreements, *IDEI Working Paper*, n. 682, August 31, 2011, revised April 30, 2013

Nordhaus, W.D. and Z. Yang (1996) A regional dynamic general-equilibrium model of alternative climate-change strategies, *American Economic Review*, 86, p. 741-765

Sterner, T. and L.H. Isaksson (2006) Refunded emission payments – a hybrid instrument with some attractive properties. *Ecological Economics, 57*, p. 93-106

A Appendix

Proof of Lemma 1. For $k > 0$, implicit differentiation of (9) yields (after rearranging):

$$\frac{da^*}{dk} = \frac{b - \mu c'(a^*)h}{[1 - \mu(N - k)]c''(a^*)h}. \tag{25}$$

The denominator is positive because $1 > \mu(N - k)$ and $c''(a^*) > 0$ for any $a^* > 0$ by assumption. Moreover, using (9) to substitute for $c'(a^*)h$, the numerator becomes

$$b - \mu c'(a^*)h = \frac{b(1 - \mu N)}{1 - \mu(N - k)} > 0$$

since $\mu N < 1$ by assumption. □

Proof of Proposition 1. Define $\Delta\Pi(k, a) \equiv \Pi_s(k, a) - \Pi_n(k, a)$. Substitution from (7) and (8), thereby using (1), yields at the equilibrium abatement target $a^*(k)$,

$$\Delta\Pi(k, a^*(k)) = c'(a^*(k))h\mu N\left[\bar{E} - a^*(k)\right] - (1 - \mu N)c(a^*(k))h. \tag{26}$$

Differentiation w.r.t. k and again simplifying terms gives

$$\frac{d\Delta\Pi(k, a^*(k))}{dk} = \left[c''(a^*)\mu N(\bar{E} - a^*) - c'(a^*)\right] h\frac{da^*}{dk}. \tag{27}$$

From (9) and the curvature assumptions, $a^*(k) = 0$ at $k = 0$, and $a^*(k) > 0$ for all $k > 0$. Hence da^*/dk and – again using the curvature assumptions – the whole term (27) are non-negative at $k = 0$. Moreover, for $k > 0$, substituting for da^*/dk from (25) gives

$$\frac{d\Delta\Pi(k, a^*(k))}{dk} = \left(\mu N(\bar{E} - a^*) - \frac{c'(a^*)}{c''(a^*)}\right)\frac{b - \mu c'(a^*)h}{1 - \mu(N - k)}. \tag{28}$$

The second term is always positive from the proof of Lemma 1. Moreover, $\mu Na + c'/c''$ increasing in a is a sufficient condition for the first term to be decreasing in $a^*(k)$ and, therefore, in k (by Lemma 1). Accordingly, $d\Delta\Pi/dk$ is non-negative at $k = 0$ and can change its sign at most once thereafter (from positive to negative) if $\mu Na + c'/c''$ is (weakly) increasing in a. Hence, as $a^*(0) = 0$ and $\Delta\Pi(0, 0) = 0$, there can be at most one solution $k > 0$ that satisfies $\Delta\Pi(k, a^*(k)) = 0$, and at this solution $d\Delta\Pi/dk < 0$, i.e. it satisfies (12). □

Proof of Proposition 2. Suppose to the contrary of the proposition that there exists an integer $k' < I(k^*)$ that satisfies inequalities (10), in particular the external stability condition (for ease of notation, we write $\Pi_i(k)$ for $\Pi_i(k, a^*(k))$)

$$\Pi_n(k') \geq \Pi_s(k' + 1). \tag{29}$$

Recall that $d\Delta\Pi/dk$ is non-negative at $k = 0$ (see the proof of Proposition 1) and $\Delta\Pi(0) = 0$. If k^* is the unique solution to (11), then it also satisfies (12) (by Proposition 1). It follows that $\Pi_n(k) \leq \Pi_s(k)$ for all values of $k < k^*$.[30] Moreover, since $I(k^*)$ is the largest integer smaller than or equal to k^* and $k' < I(k^*)$, we have $k' < k'+1 \leq I(k^*) \leq k^* \leq N$. Therefore, $\Pi_n(k') \leq \Pi_s(k')$ and $\Pi_n(k'+1) \leq \Pi_s(k'+1)$. Finally, the overall surplus is clearly monotonic in k since this implies a larger internalization of the externality and BCA effects cancel out in the aggregate, i.e., $k'\Pi_s(k') + (N-k')\Pi_n(k') < (k'+1)\Pi_s(k'+1) + (N-(k'+1))\Pi_n(k'+1)$. Collecting the above yields

$$\begin{aligned} &N\Pi_n(k') \leq k'\Pi_s(k') + (N-k')\Pi_n(k') \\ &< (k'+1)\Pi_s(k'+1) + (N-(k'+1))\Pi_n(k'+1) \leq N\Pi_s(k'+1), \end{aligned}$$

a contradiction to (29). □

Proof of Proposition 3. Follows from the arguments in the main text. □

Proof of Proposition 4. We only need to consider interior solutions at the participation stage. For these, recall that the equilibrium values $a^*(R)$ and $k^*(R)$ are defined by (9) and (11), i.e. they solve

$$bk(R) - [1-\mu(N-k(R))]\, c'(a(R,k(R)))h(R) = 0 \tag{30}$$

$$(1-\mu N)c(a(R,k(R))) - \mu N c'(a(R,k(R)))[\bar{E} - a(R,k(R))] = 0. \tag{31}$$

These conditions must hold for all values of R; hence we can differentiate them with respect to R. Doing so for (31) yields after rearranging

$$\left[c'(a^*) - \mu N c''(a^*)(\bar{E} - a^*)\right] \frac{da^*}{dR} = 0. \tag{32}$$

From Proposition 1, $\frac{d\Delta\Pi(k,a^*(k))}{dk} < 0$ at the equilibrium values $a^*(R), k^*(R)$. From (27) this implies that the term in square brackets is positive. Hence $da^*/dR = 0$. Using this when differentiating (30) with respect to R and rearranging yields

$$\frac{dk^*}{dR} = \frac{[1-\mu(N-k^*)]c'(a^*)h'(R)}{b - \mu c'(a^*)h(R)}. \tag{33}$$

The numerator is negative by the curvature assumptions, and the denominator is positive by the proof of Lemma 1; hence $dk^*/dR < 0$.

Turning to the welfare result, differentiating expected payoffs gross of R&D costs, i.e. $\Pi(R) \equiv \pi(R) + f(R/N)$, yields (using (16), and suppressing asterisks * for ease of

[30] This is also the case for a grand coalition with $\Pi_n(N) < \Pi_s(N)$ since in this case there can be no k for which $\Pi_n(k) = \Pi_s(k)$ (because $\Delta\Pi(0) = 0$ and $\Delta\Pi(k)$ can change its sign at most once thereafter from positive to negative; see proof of Proposition 1).

notation)

$$\begin{aligned}\frac{d\Pi(R)}{dR} &= k'(R)(ba - c(a)h(R)/N) - k(R)c(a)h'(R)/N \\ &= \frac{kh'(R)}{N(1-\mu N)h(R)}\left[N(1-\mu N+\mu k)ba - (2-2\mu N+\mu k)c(a)h(R)\right],\end{aligned}$$

where the second line follows from using (33) to replace $k'(R)$, and then (30) to replace $c'(a)$. Since $h'(R) < 0$ and $1 > \mu N$, it remains to show that the expression in square brackets is positive. Using (30) to replace $h(R)$, this yields the following condition:

$$\varepsilon \equiv \frac{ac'(a)}{c(a)} > \frac{k(2-2\mu N+\mu k)}{N(1-\mu N+\mu k)^2}.$$

The right-hand side is positive so that the elasticity, ε, of the cost function $c(a)$ must be sufficiently large. Specifically, adding and subtracting μk in the numerator on the right-hand side, the expression becomes

$$\varepsilon > \frac{k}{N}\left(\frac{2(1-\mu N+\mu k)-\mu k}{(1-\mu N+\mu k)^2}\right) = \frac{k}{N[1-\mu(N-k)]}\left(2-\frac{\mu k}{1-\mu N+\mu k}\right).$$

It follows that $d\Pi(R)/dR < 0$ if and only if this is satisfied. Moreover, in this case also net welfare $\pi(R)$ is decreasing in R because $f(r)$ is increasing in $r = R/N$. □

Proof of Proposition 5. It remains to show that under a fixed participation of $k = N$, a positive R&D effort is obtained in equilibrium. This rules out the case where $R = 0$ and $k = N$. The proposition then follows directly from the result that payoffs are decreasing in R for interior solutions (Proposition 4).

To show this, note that with a fixed participation of $k = N$, the equilibrium value of R (denoted R_N) follows from choosing individual R&D investments, r_i, so as to maximize a country's expected payoff (16), respectively the sum of all countries' expected payoffs, depending on whether countries interact non-cooperatively or cooperatively at the R&D stage. The resulting first-order conditions are given by (21) and (22), respectively (i.e., further below in the text, where we analyze this problem in more detail). Moreover, equilibrium abatement satisfies the standard first-order condition (9), evaluated at $k = N$, which yields (20). Since $c'(0) = 0$, (20) clearly implies that $a > 0$. But with $a > 0$, (21) and (22) cannot be fulfilled for $R_N = 0$, because $f'(0) = 0$ and $h'(0) < 0$ by assumption. □

Proof of Proposition 6. The first two claims follow from the discussion in the main text (preceding the proposition), and using Propositions 4 and 5. It remains to show that $R_0'(\mu) > 0$, and that a higher μ makes it more likely that an interior solution $R_0 > 0$ of system (18) and (19) exists.

For interior solutions $R_0 > 0$ of system (18) and (19), implicit differentiation yields (after rearranging and again using (18) to simplify terms):

$$\frac{dR_0}{d\mu} = \frac{c(a^*)c''(a^*)h(R_0)}{\left[\mu N c''(a^*)(\bar{E} - a^*) - c'(a^*)\right] \mu c'(a^*)h'(R_0)} \text{ , and} \tag{34}$$

$$\frac{da^*}{d\mu} = -\frac{c(a^*)}{\left[\mu N c''(a^*)(\bar{E} - a^*) - c'(a^*)\right] \mu}. \tag{35}$$

The term in square brackets in the denominator of (34) and (35) is negative from Proposition 1 and equation (27) in its proof. Using this and the curvature assumptions, it follows that $R_0'(\mu) > 0$ and $a^{*\prime}(\mu) > 0$. In addition, a higher μ makes it more likely that an interior solution $R_0 > 0$ of system (18) and (19) exists. To see this, note that there is no $R_0 \geq 0$ that solves (19) if $bN > c'(a)h(0)$. Given the convexity of $c(\cdot)$, a higher a makes this case less likely, and (35) implies that a^* increases in μ. $\square$

Proof of Proposition 7. Follows from the arguments in the main text. $\square$

Discussion of the second-order condition for R_N^{nc}:
In the proof of Proposition 5 we have already shown that when $k = N$ is fixed, then countries choose positive R&D efforts (hence, $R_N^{nc} > 0$). However, the system (20) and (21) can have several solutions in the range $0 < R \leq R_0$. To see, note that the second-order condition for optimal R&D investments follows from differentiating (21) with respect to r_i, and implicit differentiation of (20) to substitute for $a'(R)$. This yields:

$$\frac{(c'(a^*)h'(R))^2}{c''(a^*)h(R)} - c(a^*)h''(R) - f''(R/N) < 0. \tag{36}$$

This condition is not generally fulfilled, and there is an intuitive explanation for that. Namely, if countries invest more in R&D, then abatement costs fall for any given abatement target. However, also the optimal abatement target of the grand coalition rises. This gives scope for increasing returns to investments in R&D, which may lead to multiple local maxima of the function $\pi(R)|_{k=N}$. Nevertheless, we can state a condition under which such a case cannot arise. In particular, dropping $f''(\cdot)$ in (36), a sufficient (albeit not necessary) condition for the above second-order condition to be fulfilled is

$$\frac{c(a)c''(a)}{(c'(a))^2} \geq \frac{(h'(R))^2}{h(R)h''(R)}.$$

Because this condition is not very intuitive, let us look at an example. Let the cost function be isoelastic (as specified in (14)). Furthermore, let $h(R) = 1/R$.[31] In this case, it is easy to verify that $c(a)c''(a)/(c'(a))^2 \geq 1/2$ when the elasticity of the cost function

[31] These functional forms are also used to construct Figure 3 in the numerical analysis of Section 4.

z is greater than or equal to 2, and $(h'(R))^2/h(R)h''(R) = 1/2$. Hence, for this example, the second-order condition for R_N to be a maximum is always fulfilled.

B Appendix for Referees

Let us provide some more details about the numerical analysis of Section 4. If embedded emissions are taxed in the importing country, the net transfers to country i are

$$t_i = (N-1)\mu p_i(\bar{E} - a_i) - \sum_{j \neq i} \mu p_j(\bar{E} - a_j). \tag{37}$$

The first term reflects country i's revenues collected from foreign firms that face the domestic carbon price p_i because their output is imported by country i. The second term reflects the carbon price paid by firms in country i that export their output to countries $j \neq i$. Note that a rise in the domestic emissions price cannot protect exporting domestic firms from tax payments in the foreign countries (see (37)). Therefore, the only strategic incentive for a country i with regards to transfers is to raise the domestic emissions price in order to collect more revenue from *foreign* firms who's output is imported. However, as discussed in Section 2.1, such a behaviour probably would violate WTO-rules. Accordingly, we extent our earlier Assumption 2 to non-signatories and assume that all countries neglect effects on BCA-induced transfers when choosing their emission reduction policy.

When there are k signatories that implement an identical emissions price p_s, and $N-k$ non-signatories that set a price p_n, transfers are (using (37)) as given by (23) and (24). Using this and applying Assumption 2 also to non-signatories, their climate policies follow from the first-order condition (using (2)):[32]

$$b = (1 - \mu(N-1))c'(a_n^*)h. \tag{38}$$

Since $p_n = c'(a_n)h$, it follows that each non-signatory implements an emissions price greater than b (the Nash equilibrium price without BCA and with fixed firm locations). Intuitively, BCA enables also non-signatories to shift part of the abatement costs to firms located in other countries. Moreover, note that p_n^* is independent of participation and abatement of the signatories (due to our assumption of linear benefits). Climate policies of signatories are the same as in the case where non-signatories set $p_n = 0$, and satisfy (9).[33] This follows straightforwardly since the two approaches differ only in the specification of transfers, which are ignored when choosing emissions (by Assumption 2).

Given the complexity of the model, general results are hard to obtain. Hence we assume that the cost function is quadratic: $c(a) = a^2/2$. This way, we can derive results

[32]It is easy to verify that the second-order condition (SOC) is fulfilled.

[33]Conditions (38) and (9) imply $a_s^*(k) > a_n^*$ for all $k > 1$.

for equilibrium abatement and participation in closed-form, for a given value of R resp. $h = h(R)$.

Using (1) and $c(a) = a^2/2$, we find the simple relation

$$p_i = a_i h,$$

which can be used to replace p_i in the expression for transfers, (37). Applying the first-order conditions (38) and (9) for a_s and a_n, and inserting the results back into the welfare functions (3) and (4) (gross of R&D investment costs), we can solve the equilibrium condition at the participation stage (11) for k to obtain:

$$k^* = \frac{(1-\mu N)(b(1+\mu N) - 2h\bar{E}\mu N(1-\mu(N-1))}{2h\bar{E}\mu^2 N(1-\mu(N-1)) - b(1+2\mu-\mu^2 N(N-2))}.$$

This condition is also used to determine the relevant range of values of h for which participation lies between 1 and N (see Figure 3).

In the next step, we insert the above value of k^* back into the gross welfare function to obtain the resulting welfare per country in case of an interior solution ($k \in (1, N)$) at the participation stage. The resulting expression for welfare (not shown) does not permit a general conclusion regarding the dependency of Π from h. Hence the analysis in Section 4 is restricted to numerical simulations. In particular, the U-shape of the gross welfare function illustrated in Figure 3 (for a specific set of parameter values) is very robust, and obtained also for large variations of the underlying parameter values (we obtained it for all variations that we considered). Note also that the U-shape does not depend on the specification of the function $h(\cdot)$. The specification $h(R) = 1/R$ was chosen only for illustrative purposes; gross welfare can be plotted directly over h, so that no specification of this function is needed to obtain the result.