ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

The political economy of negotiating international carbon markets

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Arvaniti M., Habla W. (2021). The political economy of negotiating international carbon markets. JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT, 110, 1-24 [10.1016/j.jeem.2021.102521].

Availability:
This version is available at: https://hdl.handle.net/11585/855095 since: 2022-02-10

Published:

DOI: http://doi.org/10.1016/j.jeem.2021.102521

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

10 October 2026

This is the final peer-reviewed accepted manuscript of:

Arvaniti, M., & Habla, W. (2021). The political economy of negotiating international carbon markets. *Journal of Environmental Economics and Management*, *110*, 102521.

The final published version is available online at:

https://doi.org/10.1016/j.jeem.2021.102521

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

The Political Economy of Negotiating International Carbon Markets *

Maria Arvaniti[§] and Wolfgang Habla[†]

[§] Corresponding Author: Centre of Economic Research, ETH Zurich
Zürichbergstrasse 18, 8092 Zürich, Switzerland; marvaniti@ethz.ch
[†] Department of Environmental and Resource Economics, Environmental Management,
ZEW – Leibniz Centre for European Economic Research
L7, 1, 68161 Mannheim, Germany; habla@zew.de

Abstract: International carbon markets are a cost efficient instrument for achieving a given CO_2 emissions target. This paper identifies the conditions under which participating governments benefit from the establishment of such a market through Coasean cooperation, in the presence of strategic considerations in the negotiations. While negotiations internalize cross-border spillovers, governments have incentives to send delegates who value environmental damages less than they do themselves. We find – for benefit and damage functions typically assumed in the literature on international environmental agreements – that even though total emissions are lower than in a non-cooperative regime, high cost savings from trading permits are not sufficient for all participating governments to benefit from an international market. An international market constitutes a Pareto improvement in particular if i) marginal damages substantially differ across countries, ii) marginal damages *and* marginal abatement costs are sufficiently similar across countries. Bilateral carbon markets among the large emitters China, the EU and the US would not be Pareto improving. What is more, bilateral agreements with non-tradable emissions caps may be better suited to establish mutually beneficial cooperation in such cases. We conclude our analysis with a robustness check for different functional forms.

Keywords: cooperative climate policy, political economy, emissions trading, linking of permit markets, strategic delegation, strategic voting

JEL-Classification: D72, H23, H41, Q54, Q58

* Arvaniti acknowledges the generous financial support of the Tore Browaldh Foundation (Handelsbanken foundation). Habla acknowledges financial support of the DFG Collaborative Research Center SFB 884 "Political Economy of Reforms" (project B10) at the University of Mannheim and from MISTRA via the program Carbon Exit. Declarations of interest: None

1 Introduction

Economic theory suggests that the provision of a transboundary public good, such as a stable climate on Earth, increases when decisions are taken cooperatively by governments, since cooperation internalizes cross-border spillovers. As such, international negotiations hold the promise of making every participating party better off as compared to non-cooperation, but they are also prone to manipulation: a participating government might appoint a delegate with a certain well-known political agenda that is less ambitious than it would have been in the absence of cooperation. If this agenda can credibly signal that the delegate's country has not much to gain from the negotiations, the other parties involved in the negotiations have to step up their contributions, which is beneficial for the considered country.[1] In the case of climate change, this would imply sending a less "green" delegate to the negotiations.[2] As all countries face the same incentives, they are likely to end up in a prisoners' dilemma, and it is unclear whether negotiations are still beneficial for every party. If they are not, there are strong incentives to withdraw from the agreement at a later point, or to not enter the negotiations at all.

In the literature, the mechanism described above is referred to as "strategic delegation" or "strategic voting", depending on who selects the delegate, i.e., whether the principal on whose behalf the negotiations are conducted is an institution such as the government or the electorate (the median voter, to be more specific). In this paper, we investigate how the combination of cooperation and strategic delegation affects the prospects of establishing a particular type of international environmental agreement (IEA), namely an international emissions permit market. Such markets have been proposed as one potential instrument for the efficient reduction of GHG emissions (Flachsland et al., 2009; Jaffe et al., 2009; Green et al., 2014). The obvious gain from such markets is the equalization of marginal abatement costs across firms and countries, which is a necessary condition for efficiency (Montgomery, 1972). This efficiency gain is higher the more marginal abatement costs diverge across countries

[1] As Perino (2010) shows, delegation can also serve as a commitment device when governments face time-inconsistency.

[2] Two of President Donald Trump's most controversial nominees for environment posts—one a coal industry lobbyist, the other a former Texas environmental official who dubbed carbon dioxide the "gas of life"—can be seen as signaling his views over the climate change issue and his commitment to a business friendly environmental policy.

in the absence of trading. At the same time, these markets necessarily involve a transfer payment from one country to the other (unless countries have the same marginal abatement costs before trading permits), and hence countries negotiating an international carbon market have an interest in either increasing or decreasing this transfer. This feature adds another strategic dimension to the negotiations.[3]

We model the hierarchical structure of delegation and cooperation as follows. At the first stage of the game, the principals of two countries play a Nash game by simultaneously selecting an agent each from the continuum of available agents in their economies. We find that they have an incentive to appoint agents that care less about environmental damages than they do themselves – a result that is well known in the strategic voting and strategic delegation literature and is due to the strategic substitutability of policies (Segendorff, 1998; Siqueira, 2003; Buchholz et al., 2005; Graziosi, 2009; Habla and Winkler, 2018). In particular, the principal that suffers less from higher aggregate emissions has very strong incentives to misrepresent her preferences by selecting a delegate who puts little or even no weight on climate damages. By doing so, the principal credibly commits to a less ambitious climate policy, passing on the burden of abatement to the other principal. In the second stage, the appointed agents negotiate on the total number of emission allowances and their distribution. If they fail to agree, they set up domestic policies (either domestic permit markets or a domestic tax).[4] We model the negotiations using the Nash Bargaining Solution. The negotiated policies are jointly efficient from the appointed agents' point-of-view. In the final stage, emission permits are traded, either on domestic markets or on an international market, depending on whether the negotiations were successful.

The contribution of this paper is threefold. First, we show – for benefit and cost func-

[3] In fact, even without strategic delegation, Helm (2003) shows that in the absence of a central authority determining the initial allocation of emission permits, the mere possibility of trading creates incentives for an over-provision of permits and a potential increase in aggregate emissions as compared to a situation without a permit market. In this paper, we assume a cooperative framework for the choice of permits while still allowing for a non-cooperative choice of delegates by the governments.

[4] Our main results hinge on the assumption that the principals leave the decision power with the appointed agents in case negotiations fail. This type of delegation is referred to in the literature as "strong delegation" (Segendorff, 1998) and stands in contrast to "weak delegation", where the principals rescind the agents' decision power and decide themselves, rather, when negotiations have failed. Under weak delegation, negotiations are by definition always a Pareto improvement, as the principals can only gain through the negotiations relative to the outside option, which is determined by themselves.

tions typically assumed in the literature on international environmental agreements – that cooperation in combination with strategic delegation leads to lower aggregate emissions than a non-cooperative regime. That is, the costs of strategic delegation do not offset the benefits of cooperation in terms of aggregate emissions. From an environmental perspective, this is good news.

Second, despite lower aggregate emissions and thus lower individual damages, we find that cooperation in the form of an international permit market may not be a Pareto improvement for the principals of the negotiating countries (governments or median voters); rather, cooperation may be detrimental to one of the principals relative to a non-cooperative regime. This is due to the strategic delegation incentives that the principals face, and due to the corresponding transfers, which may be quite high. We identify four cases for which negotiations about an international carbon market are mutually beneficial for the principals of two countries, whereby: i) the principals' marginal damage costs are identical and marginal abatement costs (for given abatement levels) are sufficiently similar; ii) countries are identical in terms of marginal abatement costs (for given abatement levels) *and* marginal damages as perceived by the principals are either sufficiently similar or sufficiently dissimilar; iii) countries are identical or very similar in terms of marginal abatement costs (for given abatement levels) *and* marginal damages; iv) the principals' marginal damage costs are sufficiently different. Our results imply that, in contrast to what one would expect, strongly diverging marginal abatement costs across countries (for given abatement levels), which yield high efficiency gains under permit trading, are not sufficient for the principals to establish an international permit market through Coasean cooperation in the presence of strategic delegation. We also show that bilateral trade among the big emitters EU, US and China, does not constitute a Pareto improvement, as they are very similar along one dimension (marginal damages) but not sufficiently similar with respect to the other dimension (marginal abatement costs).

Third, negotiating an IEA without transfers, i.e., an agreement with *non-tradable* emissions permits is – for empirically relevant combinations of marginal benefits and marginal costs – a more promising way to establish mutually beneficial cooperation and achieve emissions reductions than negotiating an agreement with tradable caps. The reason for this result is that the possibility of receiving a transfer under an international permit market is being exploited by the principals, as they tend to strategically appoint even less green agents than in the absence of transfers. A high transfer and

high aggregate emissions due to this behavior eventually make it unattractive for one principal to cooperate via an international permit market. This is less likely to happen when transfers between countries are not possible.[5] Among the EU, the US and China, bilateral agreements with non-tradable emissions would indeed be Pareto-improving.

Our analysis deviates from previous papers in the strategic delegation and strategic voting literature in an important aspect. While it is conceivable that for traditional public goods the costs of provision are linear and the benefits are concave, the opposite is true in the case of the climate as a public good: the costs of providing this good are convex and the benefits from this good are (approximately) linear (we justify this assumption in Section 3 but also explore the implications of other functional forms in Section 8). This change in the curvature of cost and benefit functions induces an important difference. With linear costs, the country that is more efficient at providing the public good will produce all of it and is compensated for its efforts by the other country in the Nash Bargaining Solution. In the case of climate change, however, it is arguably never optimal that one country provides all of the public good, i.e., abates all emissions it has. In this context, we show that asymmetries between countries in terms of marginal costs and marginal benefits do matter for the success of an agreement (and not only the curvature of the demand function for the public good, as found by Loeper, 2017). Moreover, assuming a quadratic abatement cost function guarantees that even for symmetric countries, there exists an equilibrium in which both agents have a positive valuation for the environment, with total emissions limited to no-policy emissions. This result is in sharp contrast with Buchholz et al. (2005) who find that with perfect transboundary environmental spillovers, both principals choose agents with a zero valuation for environmental damages, resulting in aggregate emissions soaring to plus infinity. All of these aspects show that the analysis of international agreements in the context of climate policy may substantially differ from and lead to other conclusions than analyses of more traditional transboundary public goods problems.

[5] By contrast, we find only for very few parameter combinations that the cooperative regime without transfers is a Pareto improvement for the principals compared to the cooperative regime with transfers but we also find only for few parameter combinations that the opposite holds.

2 Literature

Our paper contributes to several strands of literature. In assuming that countries are not represented by one welfare-maximizing decision maker, we explicitly account for the principal-agent relationship between different bodies involved in international policy making within a single country – for example, an incumbent government or president that serves as the principal, and a selected executive or government agency that serves as an agent. In this regard, we heavily draw on the strategic delegation literature pioneered by Schelling (1980), Vickers (1985), Fershtman and Judd (1987), and Sklivas (1987), and the strategic voting literature (Persson and Tabellini, 1992).

Siqueira (2003), Buchholz et al. (2005), Roelfsema (2007) and Hattori (2010) analyze strategic voting in the context of environmental policy.[6] While the first three contributions exclusively focus on environmental taxation, Hattori (2010) also examines the outcome of strategic voting under emissions caps. Siqueira (2003) and Buchholz et al. (2005) both find that voters' selection of agents is biased toward politicians who are less green than the median voter. By electing a more "conservative" politician, the home country commits itself to a lower tax on pollution, shifting the burden of a cleaner environment to the foreign country. In contrast, Roelfsema (2007) accounts for emissions leakage through shifts in production and finds that median voters may delegate to politicians who place greater weight on environmental damage than they do themselves whenever their preferences for the environment, relative to their valuation of firms' profits, are sufficiently strong. However, this result breaks down in the case of perfect pollution spillovers, such as the emission and diffusion of greenhouse gases. Hattori (2010) allows for different degrees of product differentiation and alternative modes of competition, i.e., competition on quantities, but also on prices. His general finding is that, when the policy choices are strategic substitutes (complements), a less (more) green policy maker is elected in the non-cooperative equilibrium. Lange and Schwirplies (2017) analyze the strategic delegation incentives in international climate negotiations when agents are concerned about the distribution of the abatement burden. Our work is also closely related to Habla and Winkler (2018), who consider non-cooperative policies. In contrast, we study the case in which total emissions and country-specific permits are decided on a centralized level, employing a Nash Bargain-

[6] Strategic delegation in the provision of public goods other than the environment is examined by, e.g., Harstad (2010), Christiansen (2013) and Kempf and Rossignol (2013).

ing Solution. We thus explore whether cooperation makes linking to an international permit market more attractive.

In the context of cooperative policies, Loeper (2017) analyzes the provision of public goods with cross-border externalities by representative democracies. Loeper finds that once voters' incentives are taken into account, whether cooperation is beneficial depends neither on voters' preferences, nor on the magnitude of spillovers, nor on the size, bargaining power and efficiency of each country. Instead, it depends only on the curvature of the demand for the public good: cooperation increases (decreases) public good provision when the demand function is more (less) convex than the unit elastic demand function. Hence, the desirability of international cooperation depends mostly on the type of public good considered. In contrast to this, we find that whether a principal benefits from cooperation or not depends on the characteristics of the countries participating and in particular on the marginal benefit and marginal cost parameters. In line with Loeper (2017), we find that allowing for transfers across countries can make cooperation detrimental.

We further contribute to the literature that asks whether and under which conditions the linking of emissions trading schemes is in the best interest of each individual country. Babiker et al. (2004) show in a partial equilibrium model and a calibrated computable general equilibrium (CGE) model that linking leads to higher social costs if the permit price interacts with distortionary domestic taxes. Marschinski et al. (2012) analyze linkage in a general equilibrium model and identify a terms-of-trade effect, which may lead to a deterioration of welfare under an international permit market. Anger (2008) shows in a two-sector general equilibrium model that linkage may not be beneficial if only one sector is linked and the national allocation of allowances towards the two sectors is endogenous. Doda and Taschini (2017) argue that fixed set up costs associated with linking may outweigh the efficiency gains from trade. Using a CGE model, Gavard et al. (2016) show that the limited trading of emissions permits between developed and developing countries or regions can be beneficial for all regions. Doda et al. (2019) quantify the efficiency gains from linking which accrue to an individual jurisdiction participating in an arbitrary linkage group. They also identify two independent sources of efficiency gains, namely effort- and risk-sharing gains. We add political economy aspects to this literature by modeling the hierarchical structure of linkage decisions and allowance choices and show that such considerations may well be a reason for the rejection of otherwise beneficial policies.

Finally, with respect to hierarchical policy structures within countries, our paper is related to Habla and Winkler (2013) and Marchiori et al. (2017), who analyze the influence of legislative lobbying on the formation of international permit markets and international environmental agreements, respectively.

3 The model

We consider two (possibly heterogeneous) countries, indexed by $i \in \{1,2\}$ and $-i \in \{1,2\}$, $i \neq -i$.[7] In each country i, emissions e_i imply strictly increasing and concave country-specific benefits from the productive activities of a representative firm, $B_i(e_i)$, while global emissions $E = e_1 + e_2$ cause strictly increasing country-specific damages, $D_i(E)$. Both functions are twice continuously differentiable, and $B_i(0) = 0$ and $D_i(0) = 0$. We assume the following functional forms for benefits and damages:

$$B_i(e_i) = \frac{1}{\phi_i} e_i(\epsilon_i - \frac{1}{2} e_i) \ , \quad B_i'(e_i) = \frac{\epsilon_i - e_i}{\phi_i} \ , \quad B_i''(e_i) = B_i'' = -\frac{1}{\phi_i} \ , \tag{1}$$

$$D_i(E) = \delta_i E \ , \quad D_i'(E) = D_i' = \delta_i \ , \quad D_i''(E) = D_i'' = 0 \ , \tag{2}$$

where $\epsilon_i, \delta_i, \phi_i > 0$. The benefit function can be interpreted as a production function, with emissions as the only input. It needs to be strictly increasing. Therefore, we restrict it to the domain $e_i \in [0, \epsilon_i]$. The parameter $\epsilon_i \geq e_i$ denotes emissions in the absence of any climate policy (referred to as "no-policy emissions" henceforth), and ϕ_i is a measure of carbon efficiency, i.e., of how emissions translate into output (a higher ϕ_i implies lower carbon efficiency). For later reference, we make the following definition in order to relate ϕ_i to marginal abatement costs:[8]

Definition 1 (Marginal abatement costs)
Country i has higher marginal abatement costs than country $-i$ for the same *level of abatement if $\phi_i < \phi_{-i}$. This holds for any level of abatement, and thus the marginal abatement cost curve is steeper in i than in $-i$.*

[7] The model can be extended to $n > 2$ countries, although it would lose analytical tractability. It is also beyond the scope of this paper to analyze more complex relationships between countries, such as coalition formation. For an analysis of coalition formation in a strategic delegation framework, see Spycher and Winkler (2019).

[8] Note that abatement costs are $\text{AC} = B_i(\epsilon_i) - B_i(e_i) = (\epsilon_i - e_i)^2/(2\phi_i) = a_i^2/(2\phi_i) \equiv \text{AC}(a_i)$, where $a_i = \epsilon_i - e_i$ is the amount of abatement.

This definition shall apply whenever we compare marginal abatement costs across countries. In addition, we will employ the following substitutions: $\epsilon \equiv \epsilon_i + \epsilon_{-i}$ and $\phi \equiv \phi_i + \phi_{-i}$. We only resort to these functional forms where necessary and keep to the more general notation elsewhere.

The above functional form assumptions allow for analytical tractability, particularly for asymmetric countries, and highlight the mechanism underlying our results. Moreover, they are common in the literature on international environmental agreements. For instance, quadratic abatement cost functions are assumed in Gersbach and Winkler (2011), Harstad (2016) and Holtsmark and Weitzman (2020). The linear damage specification rules out one source of strategic interaction among countries in the model and is in line with the assumptions of complex integrated assessment and general equilibrium climate-economy models (see, e.g., Nordhaus and Boyer, 2000; Golosov et al., 2014; Gerlagh and Liski, 2018) in which climate damage is approximately linear in the greenhouse gas concentration in the atmosphere. This is because, typically, temperature is assumed to increase logarithmically with concentrations, whereas damage is assumed to be exponential or polynomial in temperature. Holtsmark and Weitzman (2020) point out that the linear damage specification is reasonable, as it is in fact the *stock* of accumulated GHG emissions that causes climate damage. The relatively small flow of emissions within a four- to ten-year period thus has an effectively linear impact on the overall stock of atmospheric GHG emissions. The fact that we consider two countries that negotiate an international carbon market, also speaks in favor of this assumption, as their impact on the stock of GHG emissions is limited. Overall, we believe that our assumptions are well suited to capture the relatively short time periods of four to ten years that carbon markets typically span (due to, e.g., problems of governments to commit themselves to policies beyond this time horizon) but we explore the implications of other types of benefit (and thus abatement cost) and damage functions in Section 8.

3.1 International climate policy

We assume that the two countries negotiate an international climate agreement. More specifically, they negotiate the parameters of an international permit market, i.e., the total number of emission permits and their distribution across the two countries. The

countries' outside options (or threat points) in the negotiations are (unlinked) national permit markets. In our setting, these are equivalent to domestic emissions taxes.

The number of permits issued to the representative domestic firm in country i amounts to ω_i.[9] As firms in both countries require emission permits in an amount equal to the emissions e_i they produce, global emissions are given by the sum of emission permits issued, $E = \omega_i + \omega_{-i}$. Restricting emissions imposes a compliance cost on the representative firms and thus reduces profits. If permits are traded internationally (in the case of successful negotiations), a firm can generate additional profits by selling permits to the firm in the other country, or reduce the compliance cost via buying permits from abroad. Thus, the profits of the representative firm in country i read:

$$\pi_i(e_i) = B_i(e_i) + p \times (\omega_i - e_i) \ , \qquad i = 1, 2 \ , \tag{3}$$

where p is the price of permits on an international market. If negotiations fail, domestic permit markets are established and $\omega_i = e_i$ holds in equilibrium, implying that the second term in the above equation vanishes. For the later analysis, we define $T_i = p \times (\omega_i - e_i)$ to be the (sign-unconstrained) financial transfer through the permit market.

3.2 Agency structure and timing of the game

In each country i, there is a principal whose utility is given by:

$$V_i = \pi_i(e_i) - \theta_i^P D_i(E) \ . \tag{4}$$

Without loss of generality, we normalize θ_i^P to unity. In addition to the principal, there is a continuum of agents of mass one in each country, whose utilities are given by:

$$W_i = \pi_i(e_i) - \theta_i D_i(E) \ , \tag{5}$$

where θ_i is a preference parameter that is continuously distributed on the bounded interval $[0, \theta_i^{\max}]$. To ensure that, in both countries, the principal's preferences are rep-

[9] The method of allocating permits has no bearing on our results. Grandfathering the permits and auctioning them are equivalent in our setting.

resented in the continuum of agents' preferences, we impose $\theta_i^{\max} > 1$. In each country, all agents and the principal thus have equal stakes in the profits of the domestic firm but differ with respect to how much they suffer from environmental damage. This may be either because damages are heterogeneously distributed, or because the monetary valuation of homogenous physical environmental damage differs.[10] We assume that all individuals (principals *and* agents) maximize their respective utilities, i.e., the principal in country i chooses *her* actions to maximize V_i, while an agent in country i makes decisions to maximize *his* utility W_i. Importantly, the preference parameters of all individuals are assumed to be common knowledge. Thus, we abstract from all issues related to asymmetric information.[11]

We model the hierarchical structure of climate policy in the following way:[12]

1. Delegation stage (agent appointment game):
   Principals in both countries simultaneously select an agent each.

2. Policy-making stage:
   The selected agents choose the total number of permits and the allocation of permits across the two countries through (utilitarian) Nash bargaining. If negotiations fail (which they never will in our setting), the agents act non-cooperatively in determining the number of permits issued in their countries, for their domestic permit markets.

3. Permit-trading stage:
   Depending on the established regime, emission permits are traded on perfectly competitive domestic markets *or* an international permit market.

[10] In the case of climate change, it does not seem unrealistic that some individuals benefit from global warming, while the majority are actually harmed, implying $\theta_i^{\min} < 0$. For instance, some individuals might perceive warmer temperatures as beneficial, while others actually suffer because they see many ecosystems and landscapes deteriorate and would prefer a more stable climate. Varying preferences with respect to global warming can also be rooted in economic gains or losses in sectors that are sensitive to a change in temperature, such as agriculture. While this remains a possibility, we focus on the case with $\theta_i^{\min} = 0$, which is the standard assumption in the literature.

[11] This may seem restrictive at first glance, but it is not in the context of our model framework. One principal's incentive to strategically delegate to an agent stems exclusively from the other principal's ability to observe the foreign agent's preferences. Moreover, high-level political delegates generally have well-known political agendas; therefore, this assumption seems to be a good description of reality.

[12] There may also be a ratification stage in which the government, the national parliament or another institution have a say over whether the outcome of the negotiations is accepted or not. As shown by Graziosi (2009), adding such a stage would limit the extent of strategic delegation.

Superscript	Description
C	Cooperative regime featuring an international permit market (with delegation)
C, NT	Cooperative regime featuring unlinked domestic permit markets (with delegation)
NC	Non-cooperative regime featuring unlinked domestic permit markets (no delegation)
D	Unlinked domestic permit markets (under the authority of the selected agents)

Table 1: The different regimes analyzed in the paper

Despite being highly stylized, the model captures essential characteristics of the hierarchical structure of international environmental policy and international negotiations. It is compatible with various delegation mechanisms that are present in modern democratic societies. For example, the principal might be the median voter among the electorate, while the agent represents the elected government. Alternatively, the principal might be the parliament or government that delegates a decision to an agent (e.g., to the minister of the environment). We prefer the latter interpretation, because it seems to be the more realistic one for an issue such as negotiations about an international carbon market.[13]

As we will make a number of comparisons in this paper, e.g., between the outcome of the regime *with* negotiations and the outcome of the regime *without* negotiations, we introduce the following notation: Superscript D stands for domestic permit markets, superscript C for cooperation featuring an international permit market, superscript C, NT for cooperation featuring domestic permit markets ("Non-Tradable" permits), and superscript NC for the regime with no international cooperation ("Non-Cooperation"). For an overview of the various regimes see Table 1.

In Section 6, we compare whether principals benefit from "delegated cooperation" or whether they would be better off if they chose policies in a purely non-cooperative fashion (comparison between regimes C and NC). This comparison can also be interpreted as an initial stage to the game outlined above. Finally, in Section 7, we examine whether negotiations over non-tradable emissions caps would result in higher or lower global emissions and individual welfare than negotiations over an international permit market (comparison between regimes C, NT and C). The former regime is equivalent to an international climate agreement in the absence of transfers between countries.

[13] While our model is cast in the latter interpretation, it is straightforward to show that our results also hold in a median voter model in which the agents constitute the voters and the principal is the politician elected under majority rule. For this, we require the preferences of the voters to be single-peaked, which is the case in our setting.

4 Delegated permit choice

Solving the game by backward induction, we first determine the equilibrium on the permit market. Profit maximization of the representative firm leads to an equalization of marginal benefits with the country-specific equilibrium permit price p_i for domestic permit markets. In the case of an international permit market, there is only one permit price, and the marginal benefits of the participating countries are equalized in equilibrium:

$$p_i(\omega_i) = B_i'(\omega_i) = \frac{\epsilon_i - \omega_i}{\phi_i} \ , \qquad i = 1, 2 \ , \quad \text{for domestic markets} \tag{6}$$

$$p(E) = B_i'\Big(e_i(E)\Big) = \frac{\epsilon_i - e_i(E)}{\phi_i} \ , \quad i = 1, 2 \ , \quad \text{for an international market} \tag{7}$$

The equilibrium permit price on the international market, $p(E) = (\epsilon - E)/\phi$, goes down as the global supply of permits increases.

Moving to the second stage of the game, we analyze the permit choices when the agents cooperate, i.e., when they have agreed on the establishment of an international carbon market. We capture the negotiation of the international agreement through the Nash Bargaining Solution (NBS) with equal bargaining weights.[14]

First, we need to examine the permit choices that the selected agents will make in case the negotiations break down, i.e., in the bargaining default (threat point), in which domestic permit markets are established. In this case, the appointed agent of country i – with preference parameter θ_i – sets the level of emission permits ω_i to maximize W_i^D (equation 5) subject to equation (6), given the permit choice ω_{-i} of the other country's agent. The reaction function of agent i is then implicitly given by:

$$B_i'(\omega_i) - \theta_i D_i' = 0 \ . \tag{8}$$

Each agent thus trades off the marginal benefits of issuing more permits against the corresponding environmental damage costs (as valued by him) in his own country.

[14] We assume equal bargaining weights for three reasons. First, it is hard to determine in reality the weight of which country is higher in the negotiations. Second, we focus on the effects that other asymmetries between countries, i.e., differences with respect to marginal damages and marginal emission benefits, have on emissions and welfare levels. Third, if countries have unequal bargaining weights, this will change the results in a straightforward way.

There exists a unique Nash equilibrium (NE):

Proposition 1 (Unique NE at stage two in the bargaining default) *For any given vector $\Theta = (\theta_i, \theta_{-i})$ of preferences of the selected agents, there exists a unique Nash equilibrium in permit choices in the bargaining default, $\Omega^D(\Theta) = \Big(\omega_i^D(\theta_i), \omega_{-i}^D(\theta_{-i})\Big)$.*

Proof. Existence and uniqueness follow from the fact the benefit benefit function $B(e_i)$ is strictly concave and twice continuously differentiable in emissions e_i. See the Appendix for closed-form solutions.

Due to the linearity of the damage function, the agents' permit choices are dominant strategies. Furthermore, as long as agent i perceives global warming as harmful, i.e., $\theta_i > 0$, he will choose an emissions level ω_i that is lower than no-policy emissions ϵ_i; otherwise, he will choose $\omega_i = \epsilon_i$.

In the negotiations, the delegated agents bargain about the total level of emissions E and the country-specific permit endowments ω_i and ω_{-i}. We denote the share of total permits allocated to countries i and $-i$ with λ and $(1-\lambda)$, respectively. Effectively, we analyze the case of an international agreement with transfers here, since one country will always be the buyer of permits while the other country will be the seller (unless the countries are perfectly symmetric). The NBS is given by the levels of E and λ that maximize the "Nash product", i.e., the product of the two agents' utilities in the negotiations in excess of their payoffs in the bargaining default:

$$\max_{E,\lambda} \quad \Big[W_i^C(\Theta) - W_i^D\Big(\Omega^D(\Theta)\Big)\Big] \times \Big[W_{-i}^C(\Theta) - W_{-i}^D\Big(\Omega^D(\Theta)\Big)\Big] \ , \tag{9}$$

where $W_i^C(\Theta)$ is agent i's welfare level in the cooperative regime C,which features a permit market (see equation 5). Let agent i's welfare gain in the cooperative scenario as compared to the default be $\Delta W_i(\Theta) \equiv W_i^C(\Theta) - W_i^D\Big(\Omega^D(\Theta)\Big)$. Then the first-order conditions yield (for notational convenience, we suppress all dependencies on Θ henceforth):

$$\begin{aligned} &\Big[p'\Big(\lambda E - e_i(E)\Big) + p(E)\lambda - \theta_i D_i'\Big]\Delta W_{-i} = \\ &\quad - \Big[p'\Big((1-\lambda)E - e_{-i}(E)\Big) + p(E)(1-\lambda) - \theta_{-i} D_{-i}'\Big]\Delta W_i \ , \end{aligned} \tag{10}$$

$$\Delta W_i = \Delta W_{-i} \ . \tag{11}$$

Substituting (10) into (11), we get the optimal levels of λ and E.

Proposition 2 (Unique optimum in the Nash Bargaining Solution) *There is a unique optimum in the Nash Bargaining Solution for any given vector* $\Theta = (\theta_i, \theta_{-i})$ *of agents' preferences, which is characterized by:*

$$p(E) - \theta_i D_i' - \theta_{-i} D_{-i}' = 0 \ , \tag{12}$$

$$\lambda = \frac{1}{2p(E)E}\Big(B_{-i}\big(e_{-i}(E)\big) - B_i\big(e_i(E)\big) + 2p(E)e_i(E) + \theta_i D_i(E) - \theta_{-i} D_{-i}(E) + W_i^D - W_{-i}^D\Big) \ . \tag{13}$$

Proof. Existence and uniqueness follow from the fact that aggregate welfare of the agents is strictly concave and differentiable in aggregate emissions E.

Equation (11) reveals that the permits are allocated in such a way that the gains from bargaining (relative to the non-cooperative solution as chosen by the appointed agents) are split equally between the two agents. This results in a share of permits for country i given by equation (13). In the absence of a permit market, an explicit monetary transfer would have to be made in order to ensure that equation (11) holds. By contrast, the permit allocation in the NBS leads to an implicit monetary transfer from one country to the other. The optimal level of emissions, E^C, which the NBS dictates, is jointly optimal from the appointed agents' points of view – it maximizes their aggregate welfare. This can be seen from equation (12), in which the marginal benefit of emissions, which is the same for the two countries and equal to the permit price, is equated with the sum of marginal damages from emissions, as perceived by the agents. This is the Samuelson condition for the optimal provision of the climate as a public good. In other words, the NBS in our framework is straightforward: the appointed agents decide a level of emissions that is jointly optimal from their perspective, and then they split the bargaining gains equally by appropriately allocating the initial permit endowment across the two countries.[15]

[15] We can see from equation (13) how the threat point, i.e., the bargaining default, affects the number of permits an agent receives and thus the transfer in the NBS: the higher is the welfare of country i's agent, W_i^D, and the lower is the welfare of the other country's agent, W_{-i}^D, at the threat point, the more permits country i is allocated in the NBS. In other words, the better off an agent is in case the negotiations fail, the more he benefits if the negotiations are successful (in terms of an increased transfer from the other country or a decreased transfer to the other country). Hence, there are strategic incentives for the principals to alter their agents' threat points, thereby

5 Strategic delegation

We now turn to the selection of agents at the first stage of the game, in which the principals anticipate the effects of their choices on subsequent stages.

The principal in country i selects an agent with preference parameter θ_i to maximize $V_i^C(\Theta)$ (equation 4), given the Nash bargaining outcome $\Omega^C(\Theta)$ at the second stage and the preference parameter θ_{-i} of the selected agent in the other country. Taking equation (7) into account, the first-order condition yields:

$$p^C(\Theta)\frac{d\omega_i^C(\Theta)}{d\theta_i} + \frac{dp^C(\Theta)}{d\theta_i}\left[\omega_i^C(\Theta) - e_i^C(\Theta)\right] - D_i'\frac{dE^C(\Theta)}{d\theta_i} = 0 \ , \tag{14}$$

which implicitly determines the reaction function of the principal of country i, $\theta_i^C(\theta_{-i})$. The terms $d\omega_i^C(\Theta)/d\theta_i$, $dp^C(\Theta)/d\theta_i$ and $dE^C(\Theta)/d\theta_i$ are given by equations (A.6b) and (A.6c) in the Appendix.

The first term in the above equation gives principal i's marginal benefit of delegating to an agent with marginally lower environmental preferences than she has herself. This benefit is equal to the number of additional permits that she receives in the negotiations multiplied with the equilibrium permit price. The second term is either positive or negative (or zero for symmetric countries), as a higher permit supply decreases the equilibrium price of permits, which is beneficial for the permit-buying country and harmful to the permit-selling country. Finally, the third term illustrates the marginal costs of strategic delegation: total emissions rise, causing additional damage to principal i.

The reaction function is downward-sloping, which implies that the choices of agents' preference parameters are strategic substitutes. There is a unique Nash equilibrium, as the following proposition states.

Proposition 3 (Unique Nash equilibrium at stage one)
There exists a unique Nash equilibrium at stage one in which the principals of both countries simultaneously choose agents, taking the other principal's choice as given.

improving their agents' bargaining positions, but there are also incentives to alter the threat point of the other country's agent (if this is possible, i.e., when $\theta_{-i} > 0$). It can easily be shown that W_i^D is higher and W_{-i}^D is (weakly) lower the lower is θ_i.

In the Nash equilibrium $\Theta^C = (\theta_i^C, \theta_{-i}^C)$, the appointed agent always has lower environmental preferences than the principal ($\theta_i^C < 1 \quad \forall i = 1, 2$). The Nash equilibrium can be of two types:

1. ***Interior NE:*** *If the principals have sufficiently similar marginal damages, i.e., $2\phi_i/(3\phi_i + 2\phi_{-i}) < \delta_i/\delta_{-i} < (3\phi_{-i} + 2\phi_i)/(2\phi_{-i})$, the appointed agents of both countries have a positive valuation of environmental damages equal to*

$$\theta_i^C = \frac{2}{3}\phi \frac{2\phi + \phi_i\left(1 - 2\frac{\delta_{-i}}{\delta_i}\right)}{2\phi_i^2 + 2\phi_{-i}^2 + 3\phi_i\phi_{-i}} > 0 \quad \forall i = 1, 2 \ . \tag{15}$$

2. ***Corner NE:*** *If the principals exhibit substantially different marginal damages, i.e., $\delta_i/\delta_{-i} \leq 2\phi_i/(3\phi_i + 2\phi_{-i})$ or $\delta_i/\delta_{-i} \geq (2\phi_i + 3\phi_{-i})/(2\phi_{-i})$, the principal with the higher marginal damage (say principal i) appoints an agent with positive valuation of damages and the other one an agent with a valuation of zero:*

$$\theta_i^C = \frac{2\phi}{2\phi + \phi_{-i}} > 0 \quad \textit{and} \quad \theta_{-i}^C = 0 \ . \tag{16}$$

Proof. See the Appendix.

The first part of this proposition states that for principals who do not differ a lot in their marginal damages, the unique Nash equilibrium is interior, with both appointed agents having a positive valuation of environmental damages. The second part states that if principal i's marginal damage is sufficiently low compared to the other principal's marginal damage, i.e., $\delta_i/\delta_{-i} \leq 2\phi_i/(3\phi_i + 2\phi_{-i})$, then she will appoint someone who has a valuation of environmental damages equal to zero. Finally, a principal will always delegate to an agent with lower environmental preferences than hers. This finding is in line with, e.g., Loeper (2017) or Buchholz et al. (2005) who also show that principals appoint agents who value the public good less than they do themselves. In contrast to the latter study, however, we show that a unique Nash equilibrium exists even in the case of symmetric countries and global pollution, and that both appointed agents have a positive valuation of environmental damages in this equilibrium.

6 When is cooperation beneficial – and for whom?

In this section, we ask whether cooperation is beneficial in terms of global emissions and individual welfare. To this end, we define the benchmark against which the outcome under cooperation is assessed to be the outcome of a Nash game where the principals establish domestic permit markets, deciding on permit issuance themselves. In other words, we compare the outcomes of the regimes C and NC. This comparison can be interpreted as an initial stage of the game, in which the principals of both countries decide whether they want to enter negotiations about a common permit market at all.[16]

In the non-cooperative regime NC, each principal maximizes $V_i^{NC} = B_i(\omega_i^{NC}) - D_i(E^{NC})$ by choosing the number of permits ω_i^{NC} issued for her domestic permit market. As we show in the Appendix, the equilibrium in the regime NC is equivalent to the equilibrium when the principals delegate to agents who choose policies non-cooperatively at the second stage (instead of negotiating an agreement) as in the threat point described in Section 4. Thus, for instance, $E^{NC} = E^D(\Theta^P)$, where $\Theta^P = (\theta_i^P, \theta_{-i}^P) = (1, 1)$ is the set of the principals' preference parameters (analogously for ω_i^{NC} and ω_{-i}^{NC}). The reason for this equivalence is that in a non-cooperative framework without an international permit market, the principals do not have an incentive to misrepresent their own preferences by selecting an agent with different preferences than their own because permit choices made by the agents are strategically neutral due to the linearity of the damage function (see Habla and Winkler, 2018, for more on this issue).

6.1 Comparison of equilibrium emissions

First, we examine whether global emissions are higher or lower in equilibrium under cooperation and delegation (regime C) than in the case when principals choose policies in a purely non-cooperative fashion, forming domestic permit markets (regime NC). We can establish the following proposition.

[16] If the principals were to negotiate the international agreement themselves, they would fully internalize all environmental externalities such that equation (12) would read $p = \delta_i + \delta_{-i}$. Clearly, the permit price would be higher and thus global emissions would be lower in this case than under the cooperative regime with delegation.

Proposition 4 (Cooperation lowers aggregate emissions)
Cooperation in the presence of delegation yields strictly lower aggregate equilibrium emissions than policies set by the principals in a non-cooperative fashion, i.e., $E^C(\Theta^C) < E^{NC} = E^D(\Theta^P)$.

Proof. See the Appendix.

For the environment, this proposition is good news, because it implies that the principals indeed achieve emissions reductions through cooperation, even though cooperation comes at the cost of strategic delegation. This proposition stands in stark contrast to the results of the study by Buchholz et al. (2005), in which damages soar to plus infinity in both the cooperative and the non-cooperative outcome in the case of global pollutants, as we assume in this paper. However, their result is an artefact of the specification of costs and benefits. Buchholz et al. assume that the benefits of emissions are linear while the costs of emissions are convex, which results in a corner solution (infinitely high emissions) in either regime.[17] As argued earlier, we assume exactly the opposite for the benefit and cost functions.

6.2 Comparison of equilibrium welfare

Despite the fact that global emissions are strictly lower under cooperation than when the principals choose policies non-cooperatively themselves, it is not clear that both principals are also better off under this regime, due to the assumed asymmetries in marginal emission benefits and marginal environmental damage costs. For example, it could be that the principal of the country that has very high marginal abatement costs compared to the other country, is better off when climate policy is less ambitious, even when she does not suffer much from environmental damage. While the gains from cooperation are shared equally between the appointed agents at stage two, no similar condition holds for the principals' welfare at stage one. In the absence of strategic considerations associated with delegation, cooperation would always be beneficial for the negotiating parties.

[17] In fact, as Loeper points out in the working paper version (Loeper, 2014) of the published paper (Loeper, 2017), the symmetric equilibrium that Buchholz et al. (2005) focus on does not exist in the case of global pollution in their framework.

Formally, a principal benefits from cooperation (under delegation) if

$$\Delta V_i = V_i^C(\Theta^C) - V_i^{NC} = V_i^C(\Theta^C) - V_i^D(\Theta^P) > 0 \ . \tag{17}$$

Furthermore, an international permit market under cooperation constitutes a Pareto improvement if equation (17) holds for both principals.

It is straightforward to show that a principal's payoff difference ΔV_i does not depend on her no-policy emissions ϵ_i (or ϵ_{-i}). Instead, only $\delta_i, \delta_{-i}, \phi_i$ and ϕ_{-i} enter condition (17). Therefore, we impose without loss of generality: $\delta_i = a\delta_{-i}$ and $\phi_i = b\phi_{-i}$, where $a, b > 0$. That is, we simulate the solution for *all* possible combinations of the parameters $\delta_i, \delta_{-i}, \phi_i$ and ϕ_{-i} without any restrictions. The parameter a then depicts the ratio of marginal damages, while b depicts the ratio of marginal abatement costs for the same levels of abatement in the two countries. We define $\Delta B_i = B_i\Big(e_i^C(\Theta^C)\Big) - B_i\Big(e_i^{NC}\Big)$ and $\Delta D_i = \Big[D_i\Big(E^C(\Theta^C)\Big) - D_i\Big(E^{NC}\Big)\Big]$ for later use.[18]

We summarize the results that can be derived analytically in the following proposition.

Proposition 5 (Mutually beneficial international cooperation)
Cooperation in the presence of delegation is mutually beneficial for the principals, i.e., $\Delta V_i > 0 \wedge \Delta V_{-i} > 0$*, in the following cases:*

1. *When* $a = 1$*: cooperation is mutually beneficial if and only if* $\phi_i^3 - \phi_{-i}^3 < 3/2\phi_i\phi_{-i}^2$ *for* $\phi_i > \phi_{-i}$ or $\phi_{-i}^3 - \phi_i^3 < 3/2\phi_i^2\phi_{-i}$ *for* $\phi_i < \phi_{-i}$ *(numerically,* $0.68 < b < 1.48$*).*
2. *When* $b = 1$*: cooperation is mutually beneficial if and only if* $a \in (0, 3/10) \cup (3/4, 4/3) \cup (10/3, +\infty)$*.*
3. *When* $a \to 0$*: cooperation is always mutually beneficial.*
4. *When* $b \to 0$*: cooperation is mutually beneficial if and only if* $a > 3$*.*

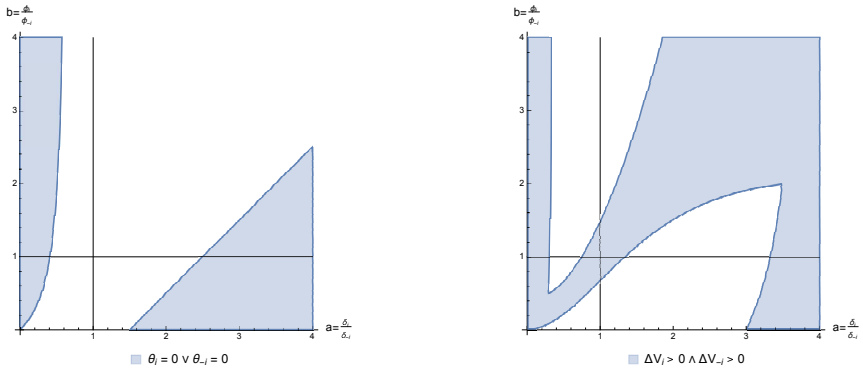
Proof. See the Appendix.

To build intuition, Figure 1 illustrates the parameter combinations of a and b for which we have a corner Nash equilibrium (and an interior Nash equilibrium, respectively) at the first stage (Figure 1a) and for which it is beneficial for both principals to enter

[18] Note that marginal abatement costs depend on the abatement levels in each country. They are equalized across countries under an international permit market. We cast the following analysis in terms of the parameters ϕ_i and ϕ_{-i}, which are, by Definition 1, a measure of marginal abatement costs in case the two countries do the same amount of abatement.

negotiations on establishing an international permit market (Figure 1b). The straight black lines depict ratios of $\delta_i/\delta_{-i} = 1$ and $\phi_i/\phi_{-i} = 1$, respectively, illustrating the knife-edge cases of either identical marginal damages among the principals or identical marginal abatement costs in both countries.[19] It is important to note that these two diagrams are exactly the same for all values of δ_{-i} and ϕ_{-i}. The reason for this is that these two parameters enter ΔV_i and ΔV_{-i} in a multiplicative way and thus only scale the principals' payoff differences but do not change their signs, and they also do not enter the values of the equilibrium preference parameters of the selected agents.[20] In the following, we discuss each of the four cases of Proposition 5 in detail.

Figure 1: Comparison of regimes C and NC in the general case



(a) Parameter space for which the Nash equilibrium is a corner solution

(b) Parameter space for which cooperation is mutually beneficial

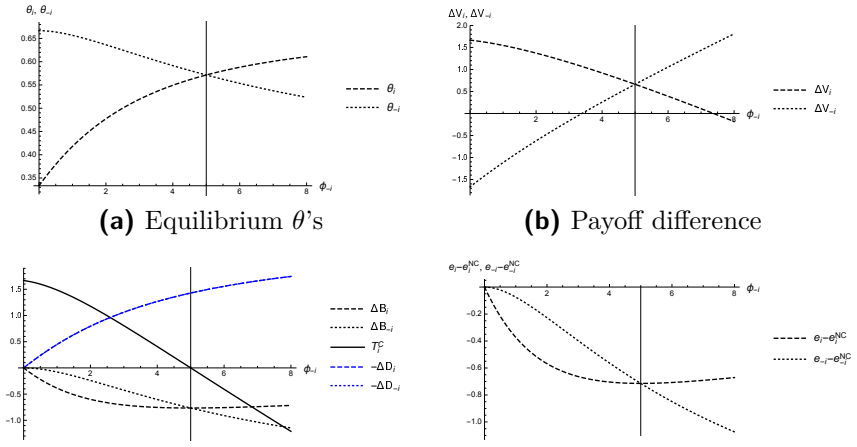
Case 1 ($a = 1$). In this case, we move along the vertical black line in Figures 1(a) and (b), and we have, by Proposition 3, a unique interior Nash equilibrium at the first stage. For cooperation to be mutually beneficial for both principals, we require not only identical marginal damages but also benefit functions that are sufficiently similar in terms of their curvature, i.e., sufficiently similar slopes of the marginal abatement costs. This result is somewhat counter-intuitive, given that the efficiency gains from trading permits, which are shared via the agreement, are higher the more marginal

[19] Note that it must hold for the two diagrams that if we reverse country indices, i.e., if we invert both a and b, we get the same result in terms of the type of Nash equilibrium and the advantageousness of cooperation.

[20] No-policy emissions are chosen such that emissions are strictly positive in all regimes for the considered ranges of a and b. In particular, $\epsilon_i = \epsilon_{-i} = 10$. We assume the same no-policy emissions in Figures 4 and 5.

abatement costs differ across countries (before trading).

Figure 2: Comparison of regimes C and NC with identical damages for both principals (assuming $\phi_i = 5$, as indicated by the vertical lines)

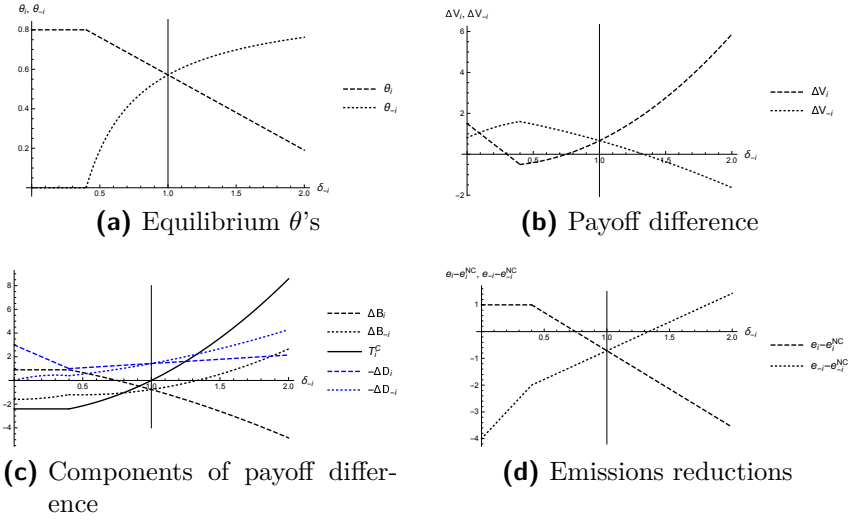


(a) Equilibrium θ's **(b)** Payoff difference

(c) Components of payoff difference **(d)** Emissions reductions

The intuition behind this result is illustrated using Figure 2, which is based on a numerical example ($\epsilon_i = \epsilon_{-i} = 10, \quad \delta_i = \delta_{-i} = 1, \quad \phi_i = 5$). The principal of the country with the higher ϕ_i, i.e., with the less steep marginal abatement cost curve, chooses an agent who is less environmentally friendly than the agent in the other country (see equation 15), as can be seen in Figure 2(a). This result is already counter-intuitive, since one would expect – in an analysis without strategic incentives – that the principal of the country with the *lower* marginal abatement costs should have a *higher* interest in pursuing more ambitious climate policy, because it is cheaper for her to do so. The explanation for why this is different with strategic incentives is that the principal with the lower marginal abatement costs has a strong incentive to downplay θ_i in order to increase the bargaining position of her agent and thus obtain a financial transfer from the other country, by receiving more permits than the country needs for covering its own emissions. In other words, this principal wants to be compensated for the cheap abatement possibilities. As the solid black line in Figure 2(c) indicates, the transfer is quite substantial for low values of ϕ_{-i} and becomes smaller the lower is the difference in marginal abatement costs between the two countries. While both principals gain from the international agreement in terms of lower aggregate emissions in the whole domain of ϕ_{-i} ($\Delta D_i > 0$ and $\Delta D_{-i} > 0$; see the blue lines in Figure 2c), they lose in terms of lower domestic production benefits (see the dotted black lines in

Figure 3: Comparison of regimes C and NC with identical curvature of benefit functions (assuming $\delta_i = 1$, as indicated by the vertical lines)



(a) Equilibrium θ's

(b) Payoff difference

(c) Components of payoff difference

(d) Emissions reductions

the same figure) due to lower aggregate emissions. Nevertheless, the lower damages under cooperation outweigh the lower benefits for both principals. However, the high transfer from country $-i$ to country i at low values of ϕ_{-i} makes it unattractive for principal $-i$ to start negotiations about a permit market in the first place. This proves that marginal abatement costs need to be sufficiently similar for identical marginal damages in order for both countries to benefit from an international permit market (see also Figure 2b).

Case 2 ($b = 1$). In this case, we move along the horizontal black line in Figures 1(a) and (b), and we have, by Proposition 3, either a unique interior Nash equilibrium or a unique corner Nash equilibrium at the first stage. We illustrate the intuition behind this result using Figure 3, which again is based on a numerical example ($\epsilon_i = \epsilon_{-i} = 15, \quad \phi_i = \phi_{-i} = 5, \quad \delta_i = 1$). As can be seen from Figure 3(a), principal i chooses a more environmentally concerned agent as long as she suffers from higher marginal damages than the principal in the other country. The reason for this is that the "high-damage" principal benefits more from emissions reductions and is thus in favor of more ambitious climate policy than the other principal, even if this comes at the expense of a financial transfer to the other country. For low values of δ_{-i}, principal $-i$ even chooses an agent with zero valuation of environmental damages. This principal would even like to choose an agent with a negative preference parameter (a case that we

exclude in our analysis). The transfer is – for low values of δ_{-i} – quite substantial and then decreases up to the point where the countries are exactly identical (see Figure 3c).

A surprising result perhaps is that both principals gain from cooperation when damages are very asymmetrically distributed. The low-damage principal (principal $-i$ to the left of the vertical line in Figure 3b and principal i to the right of this line) always benefits from cooperation, despite the lower production benefits and the relatively low environmental gain (Figure 3c). This is explained by the high transfer she receives, which also leads to principal i being worse off under cooperation for some interval of δ_{-i}. However, as the preference parameter of the appointed agent in country $-i$ hits the lower bound at zero, principal i's payoff difference between the regimes C and NC starts increasing again. This is due to the fact that at the lower bound for θ_{-i}, principal $-i$ is not able to improve the bargaining position of her agent in the negotiations anymore. In this case, i.e., for decreasing δ_{-i}, the negotiations are more successful in terms of internalizing the environmental externality than the non-cooperative outcome. In particular, aggregate emissions remain constant because of the corner solution while the regime NC leads to *higher* global emissions due to the decreasing valuation of environmental damages of principal $-i$. Relative to the non-cooperative outcome, the high-damage principal (principal i) is thus much better off under the negotiations than in the non-cooperative regime in terms of global emissions (see the increasing $-\Delta D_i$ for decreasing δ_{-i} in Figure 3c). In fact, while global emissions are lower under cooperation compared to non-cooperation, country i's emissions are higher for low values of δ_{-i} (see Figure 3d), which leads to increased domestic production benefits. To sum up, the fact that both principals gain for very asymmetric damages is caused by the corner solution.

Altogether, this implies that for both principals to benefit from cooperation, marginal damages either need to be sufficiently similar (with a ratio in the range between 3/4 and 4/3) or sufficiently different (with a ratio less than 3/10). Notice that these non-monotonicities and the irregular shape in Figure 1(b) are caused by transitions from a corner equilibrium to an interior equilibrium (or vice versa), resulting in sharp payoff changes for the principals in the cooperative regime, and/or from non-monotonicities in ΔV_i or ΔV_{-i} more generally.

From the analysis of cases 1 and 2, it becomes evident that cooperation is also mutually

beneficial when countries are completely symmetric in which case we have an interior equilibrium. The intuition for why this is the case is twofold: first, global emissions are, by Proposition 4, lower than in the non-cooperative regime; second, by virtue of symmetry, the transfer between the two countries is zero and cannot be influenced in favor of one of the principals. A third effect that is brought about by cooperation in the form of an international market is absent: there are no efficiency gains from permit trade for symmetric countries. As Figure 1(b) illustrates, cooperation is not only mutually beneficial in this knife-edge case but also as long as a and b are sufficiently similar.[21]

Case 3 ($a \to 0$). Next, we we examine the case when the ratio of marginal damages goes to zero. In this case, cooperation is always mutually beneficial. The intuition is as follows. First, when the principals' marginal damages differ a lot, then the high-damage principal has a high willingness-to-pay for emissions reductions and is thus willing to pay a high transfer to the other principal, while the low-damage principal benefits from this transfer as compared to the non-cooperative regime. Second, as we have seen earlier, it is beneficial for the high-damage principal when the low-damage principal runs into the corner solution as she cannot delegate to an agent with a negative preference parameter. The parameter combinations for which we have a corner Nash equilibrium are depicted in Figure 1(a). For low values of δ_{-i}, principal $-i$ chooses an agent with $\theta_{-i} = 0$, while for high values of δ_{-i}, principal i chooses an agent with $\theta_i = 0$. In the limit, i.e., when $a \to 0$, both principals benefit. As can be seen from Figure 1(b), this result does not only hold in the limit but already for a slightly lower than one third or slightly higher than three.

Case 4 ($b \to 0$). Finally, we examine the case when the ratio of marginal abatement costs goes to zero. In this case, we additionally require $a > 3$ for cooperation to be mutually beneficial. The intuition is again that a sufficiently large difference in the principals' willingness-to-pay for emissions reductions guarantees that the high resulting transfer leads to the emergence of cooperation.

As one can see from Figure 1(b), cooperation can also be mutually beneficial when $a, b < 1$ or $a, b > 1$, i.e., in the areas southwest and northeast of the point (1,1) in which a high-damage, low-abatement cost country meets a low-damage, high-abatement cost

[21] Note that this is not a general result. As Section 8 shows, there are functional specifications for which cooperation is not mutually beneficial in a symmetric setup.

(δ_i, c_i)	**low c_i**	**high c_i**
high δ_i	EU (13.8, 54.8) India (11.7, 65.8) China (11.0, 18.8) US (10.6, 32.9)	-
low δ_i	Russia (3.5, 93.9) Brazil (2.9, 34.6)	Japan (2.4, 219.2) Canada (1.0, 328.8) South Africa (0.7, 328.8)

Source: Holtsmark and Weitzman (2020), based on estimates from Nordhaus (2015) and McKinsey (2009).

Table 2: A high δ_i implies high marginal damages from emissions (measured in USD/tCO_2) while a high c_i implies high marginal abatement costs (for any given abatement level). For our model specification, $c_i \equiv 1/\phi_i$.

country. The intuition for this result is that the roles in this combination of countries are clearly defined, and thus strategic delegation incentives are not as strong as for other combinations of countries. For example, while one country has a high willingness-to-pay for emissions reductions and can also abate at low cost, the other country is not willing to do much and incurs a high cost of reducing emissions. Therefore, neither principal has strong incentives to shift the burden of abatement to the other country by downplaying its willingness-to-pay for emissions reductions. Instead, the principal of the high-damage, low-abatement cost country knows that it needs to do most abatement on its own, while the other principal does not care much about abatement.

Proposition 5 implies that cooperation may well be harmful to one principal due to strategic delegation. A similar result has been obtained by Loeper (2017) for the case of international negotiations over public goods. In contrast to the study by Loeper, we show that asymmetries across countries do matter in order for cooperation to be mutually beneficial.[22] More importantly, Proposition 5 demonstrates that high efficiency gains due to large ex-ante differences between countries' marginal abatement costs – the standard argument in favor of emissions trading – may not be sufficient

[22] Note that Loeper (2017) mostly examines the question of a single country being better off with cooperation and only examines Pareto improvements from cooperation for symmetric countries.

for an international permit market to be mutually beneficial.

To illustrate what these findings may imply for international environmental agreements in the real world, we provide some examples based on the analysis of Holtsmark and Weitzman (2020). Table 2 presents estimates from this study for the determining parameters for cooperation to be mutually beneficial, namely the marginal damage parameter δ_i and the carbon efficiency parameter $1/\phi_i$ for eight countries and the EU. We observe that there is a clear distinction (by one order of magnitude) between high-damage and low-damage countries, as well as between high-abatement cost and low-abatement cost countries. Interestingly, all countries (or country blocks) with high marginal damages exhibit low marginal abatement costs: the EU, India, China and the US fall into this category. Proposition 5 (case 4) suggests that bilateral cooperation among any country from the "high δ_i"-group with a country from the "low δ_i"-group, namely Japan, Canada or South Africa, would be mutually beneficial, as marginal damages differ by a factor of more than three between the two groups. On the other hand, our analysis (together with these estimates) suggests that any bilateral cooperation among China, the EU and the US would *not* be mutually beneficial.[23]

7 Cooperation with or without an international permit market?

In this section, we analyze the case when cooperation comes in the form of national carbon markets (regime C, NT), i.e, when international transfers as part of an agreement are not feasible (e.g., for political reasons). We then contrast these results with the results that we obtained when an international market is negotiated (regime C). This comparison allows us to explore the strength of the strategic dimension of an international permit market. As Harstad (2008) points out, despite the common presumption that transfers facilitate cooperation as winners can compensate losers, they also lead to the delegation of more reluctant agents, making it unclear whether side payments actually make international negotiations more successful.

Without an international market, the NBS at stage two is given by the levels of ω_i

[23] Altogether, 10 out of all 36 possible pairwise combinations of the nine considered countries/country blocks would not yield Pareto improvements under cooperation. To see this, refer to Figure 10 in the Appendix.

and ω_{-i} (which correspond to e_i and e_{-i}) that solve the following program:

$$\max_{\omega_i,\omega_{-i}} \quad \Big[W_i^{C,NT} - W_i^D\Big(\Omega^D(\Theta)\Big)\Big] \times \Big[W_{-i}^{C,NT} - W_{-i}^D\Big(\Omega^D(\Theta)\Big)\Big] \ , \tag{18}$$

where $W_i^{C,NT}$ and W_i^D are agent i's welfare levels in the cooperative regime with domestic permit markets and in the threat point (which is the same as before), respectively.

The first-order condition for the optimal $\omega_i, i = 1, 2$, is given by:

$$\Big[B_i'(\omega_i) - \theta_i D_i'\Big]\Delta W_{-i} - \theta_{-i} D_{-i}' \Delta W_i = 0 \ . \tag{19}$$

It is evident from these equations (one for every i) that, unlike for the international permit market, $\Delta W_i \neq \Delta W_{-i}$, as there is no mechanism or instrument through which gains from cooperation could be shared.

At the stage of delegation, the principal in country i selects an agent with preference parameter θ_i to maximize

$$V_i^{C,NT}(\Theta) = B_i\Big(\omega_i^{C,NT}(\Theta)\Big) - D_i\Big(E^{C,NT}(\Theta)\Big) \ , \tag{20}$$

given the Nash bargaining outcome $\Omega^{C,NT}(\Theta) = \Big(\omega_i^{C,NT}(\Theta), \omega_{-i}^{C,NT}(\Theta)\Big)$ at the second stage and the preference parameter θ_{-i} of the selected agent in the other country.

The first-order condition for principal i yields:

$$B_i'\Big(\omega_i^{C,NT}(\Theta)\Big)\frac{d\omega_i^{C,NT}(\Theta)}{d\theta_i} - D_i'\frac{dE^{C,NT}(\Theta)}{d\theta_i} = 0 \ , \tag{21}$$

where $d\omega_i^{C,NT}(\Theta)/d\theta_i$ and $dE^{C,NT}(\Theta)/d\theta_i$ are given by equations (A.29a) and (A.29c) in the Appendix.

As compared to the case with an international market in place, the first-order conditions differ in two important respects: First, marginal benefits are not equalized through the international market anymore (compare the first term in the above equation with the first one in equation 14), which would be a necessary condition to achieve efficiency (and cost-effectiveness for any given level of total emissions). Second, as there are no transfers anymore, the second term in equation (14) is missing in the

condition above. In other words, the international permit market serves two purposes: it facilitates transfers and thus makes mutually beneficial cooperation more likely; at the same time, it also ensures a cost-effective solution for given overall emissions.[24] The downside of this market is that the principals have incentives to manipulate the transfer in their favor by choosing agents who can credibly signal that their principals have a lower interest in lowering emissions than these principals actually have.

Although the system of first-order conditions given by (21) has no algebraic solution, we show in the Appendix that there exists a Nash equilibrium $\Theta^{C,NT} = (\theta_i^{C,NT}, \theta_{-i}^{C,NT})$.[25] Furthermore, while we cannot establish that the delegates' equilibrium preference parameters are always smaller than the principals' preference parameters, our numerical exercises could not generate conditions under which one of the equilibrium preference parameters would exceed unity (or equal zero).[26]

Before comparing cooperation with and without an international permit market for the general case, we first do so for perfectly symmetric countries. Although an international permit market only yields its advantages when countries are heterogeneous with respect to marginal abatement costs, it is useful to consider the symmetric setup in order to shed light on the effects of transfers through trade on the delegation outcome.

For perfectly symmetric countries, the Nash equilibrium at the delegation stage of the regime with cooperation and no international market is characterized by:

$$\theta_i^{C,NT} = \frac{3}{5} > \frac{4}{7} = \theta_i^{C} \ , \tag{22}$$

where we made use of equation (15) to find the equilibrium preference parameter for the regime with an international market.

Thus, the strategic delegation incentives are stronger for the principals (of perfectly symmetric countries) when permit trading and hence transfers between countries are possible. This finding is in line with Loeper (2017) who finds that cooperation is more likely to be beneficial without transfers. Even though transfers are zero in equilibrium

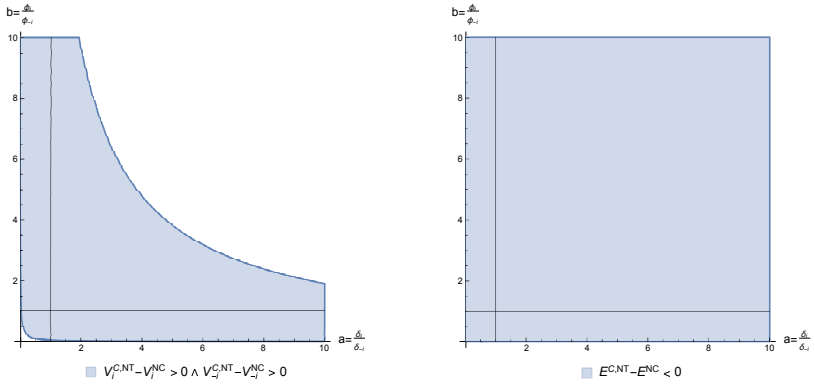
[24] Note that an internationally harmonized tax on emissions would also ensure a cost-effective solution.

[25] Numerically, we always obtain a unique Nash equilibrium for various starting values for the solution method we use.

[26] Using equations (19) and (21), we can rearrange terms to obtain: $1-\theta_i = \theta_{-i} D'_{-i} \Delta W_i / (D'_i \Delta W_{-i}) - (d\omega_{-i}^{C,NT}/d\theta_i)/(d\omega_i^{C,NT}/d\theta_i)$. While the first addend is weakly larger than zero, the second one is strictly negative. Thus, θ_i could theoretically exceed unity.

for symmetric countries, their mere existence creates extra incentives for strategic delegation, which can decrease the gains from cooperation. The reason for this is that every every principal tries to obtain a positive transfer through a higher permit allocation for her country, by selecting a less environmentally concerned agent as compared to the agent selected in the other country, which results in a "race to the bottom". Because of this, global emissions are higher and the welfare of both principals is lower under an international permit market as compared to an international agreement without trade (to see this, cf. Figure 5a and 5b). In other words, the principals of perfectly symmetric countries are definitely worse off when an international permit market is established.[27]

Figure 4: Comparison of regimes C, NT and NC



(a) Parameter space for which both principals are better off under C, NT than under NC

(b) Parameter space for which aggregate emissions are *lower* under C, NT than under NC

For the more realistic case when countries are asymmetric with respect to marginal abatement costs *and* marginal damages, Figure 4(a) depicts the parameter space for which both principals are better off under cooperation in the absence of trade compared to the non-cooperative regime.[28] As before, these two diagrams are independent of the assumed values of the parameters. This figure demonstrates that an agreement

[27] Loeper (2014, 2017) studies the effect of transfers on cooperation for the case of symmetric countries with complete spillovers but only for the two polar cases of complete and no crowding out in the policy-making stage of the non-cooperative regime. Therefore, our results are not directly comparable.

[28] As no analytical closed-form solutions exist for the regime C, NT, we first compute equilibrium values for a and b in intervals of 0.01 and then use interpolating functions to plot emissions and welfare in Figures 4 and 5. The Mathematica code can be obtained from the authors upon request.

without trade is beneficial to both principals unless countries are very asymmetric with respect to both marginal abatement costs *and* marginal damages (as perceived by the principals), i.e., in the lower left and upper right corners of the diagram. Compared to Figure 1(b), the parameter space for which cooperation is mutually beneficial is substantially larger in the absence of transfers, at least when asymmetries with respect to both dimensions are not too strong (note that the maximum of a and b for these diagrams is set to 10, whereas it was set to 4 for the diagrams of Figure 1). Furthermore, as before, cooperation *always* leads to lower aggregate emissions compared to the non-cooperative regime (see Figure 4b), also in the absence of transfers. We summarize these results as follows.

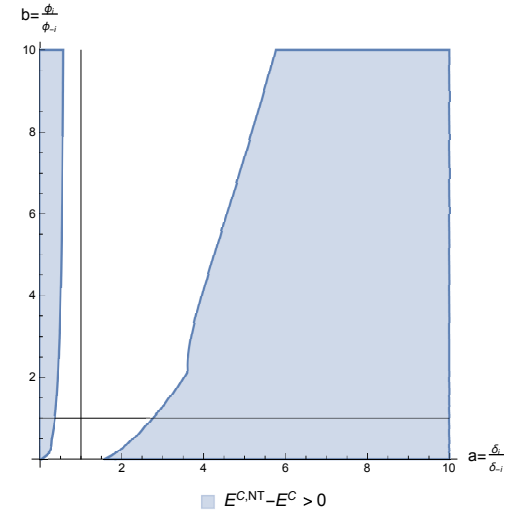
Numerical Finding 1 (Comparison of regimes C, NT and NC)
In the absence of an international permit market, cooperation in the presence of delegation always leads to lower aggregate emissions compared to the non-cooperative regime: $E^{C,NT} < E^{NC}$. Furthermore, cooperation in this form is mutually beneficial if the asymmetries with respect to marginal damages (as perceived by the principals) and marginal abatement costs are not too strong.

Coming back to our country examples from the previous section, the large emitters (EU, US, China) would now indeed be better off in all pairwise combinations of bilateral cooperation in the absence of an international market for permits, as compared to the non-cooperative regime.[29]

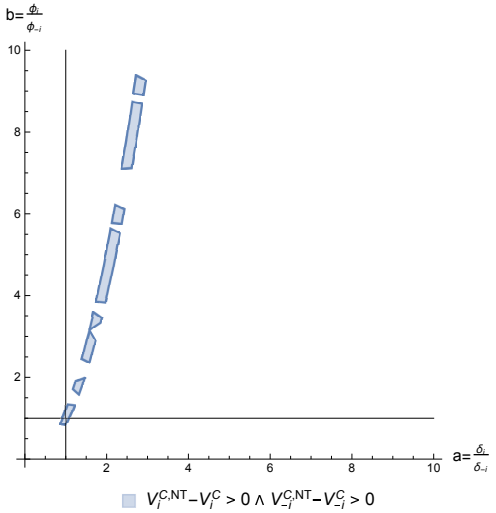
Finally, we compare aggregate emissions and payoffs of both principals under the regimes C and C, NT. As Figure 5(a) shows, aggregate emissions can either be higher or lower in the absence of trade. Interestingly, as the diagram strongly resembles Figure 1(a), it seems likely that whenever the principals run into a corner solution under an international permit market, aggregate emissions are lower under this market compared to a cooperative regime with non-tradable permits; otherwise, aggregate emissions are lower under the latter regime. Figure 5(b) shows that there exist very few parameter combinations of a and b for which both principals simultaneously achieve higher welfare in the absence of transfers than under an international permit market. However, this does not imply that the opposite is true for the remaining parameter combinations, as can be seen in Figure 5(c). Therefore, no clear picture emerges when

[29] As Figure 10(b) in the Appendix demonstrates, 13 out of all 36 pairwise combinations for these countries/country blocks are now outside of the region that depicts mutually beneficial cooperation.

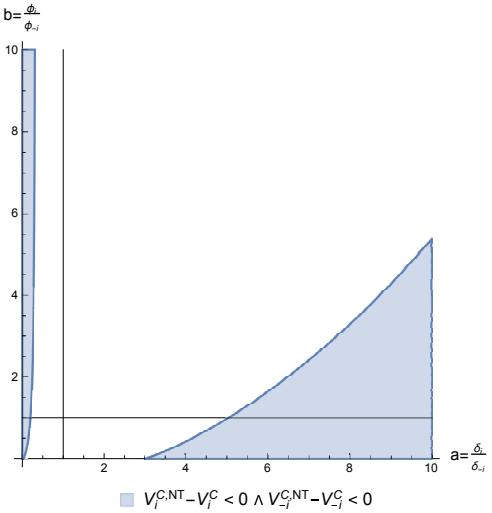
Figure 5: Comparison of regimes C, NT and C and regimes C, NT, C and NC



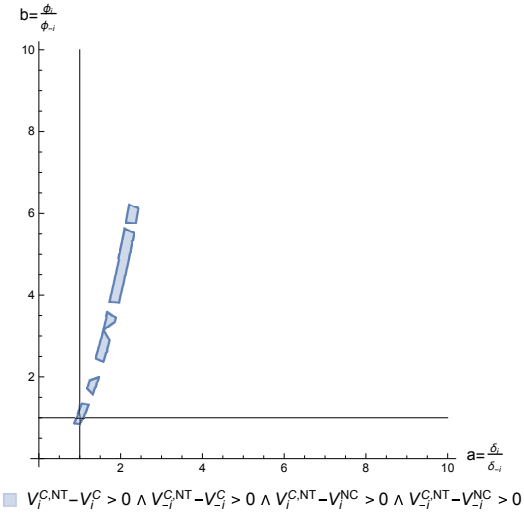
(a) Parameter space for which aggregate emissions are *higher* under C, NT than under C



(b) Parameter space for which both principals are better off under C, NT than under C



(c) Parameter space for which both principals are better off under C than under C, NT



(d) Parameter space for which both principals are better off under C, NT than under C *and* under NC

we compare cooperation with and without trade, except for the case of symmetric countries, where it is always better for both principals if there is no international market. We summarize these results as follows.

Numerical Finding 2 (Comparison of regimes C and C, NT)
Aggregate emissions may be higher or lower in regime C as compared to regime C, NT: $E^{C,NT} \gtrless E^{C}$. Furthermore, the regime C, NT is a Pareto improvement for both principals compared to regime C only for very few parameter combinations and in particular when countries are perfectly symmetric.

We can also make the "double" comparison between regimes C, NT and C and regimes C, NT and NC. Figure 5(d) depicts the parameter space for which cooperation in the absence of a carbon market is a Pareto improvement not only relative to regime C (as illustrated in Figure 5b) but also relative to regime NC (as illustrated in Figure 4a). This comparison shows that there are combinations for which cooperation in the absence of a carbon market results in Pareto improvements relative to both regimes. By contrast, there are no parameter combinations for which cooperation in the form of an international carbon market is a Pareto improvement relative to the same regimes.

In light of the above results, it is more promising in the presence of strategic delegation to negotiate an international agreement without permit trade if countries are not too asymmetric with respect to both marginal damages *and* marginal abatement costs, as this increases the parameter space for which both principals find an agreement mutually beneficial (relative to regime NC). By concluding such an agreement, the principals also achieve emissions reductions compared to a fully non-cooperative world.

8 Robustness: Other functional forms

In this section, we discuss the robustness of our results with respect to functional form assumptions related to benefits and costs of emissions. In particular, we carry out our analysis for an isoelastic and a more flexible polynomial (instead of quadratic) benefit function, and for a quadratic (instead of linear) damage function. Note that it is difficult to compare outcomes across functional forms. However, what is possible is to compare, e.g., whether cooperation still yields lower aggregate emissions and

higher individual welfare than non-cooperation for each functional form. Throughout this section, we focus on the symmetric case because closed-form solutions typically do not exist for the functional forms considered here. The same is true for the regime C, NT, which is why we only discuss comparisons of the regimes C and NC.[30]

8.1 Isoelastic benefit functions

While the quadratic benefit function (corresponding to a quadratic abatement cost function) is frequently used for the analysis of international environmental agreements, the strategic delegation literature uses various types of benefit functions (linear, isoelastic, logarithmic, or exponential functions). As Loeper (2017) demonstrates, the incentives to strategically delegate can differ significantly, depending on the curvature of the benefit function (or, equivalently, the curvature and thus elasticity of the demand function for the public good). In the following, we illustrate our results for the frequently used isoelastic specification:

$$B_i(e_i) = \begin{cases} \alpha_i e_i^{\beta_i} & \text{for} \quad 0 < \beta_i < 1 \\ \alpha_i \ln(1 + e_i) & \text{for} \quad \beta_i = 0 \end{cases} \tag{23}$$

where we assume $0 \leq \beta_i < 1$ for the benefit function to be strictly concave and $\alpha_i > 0$, with the price elasticity for the isoelastic function being equal to $\eta_i = 1/(\beta_i - 1) < 0$ and for the logarithmic function being equal to $\eta_i = -1$.[31]

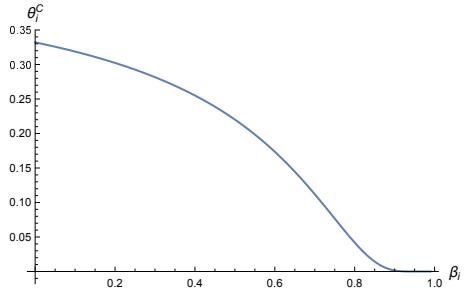
The two specifications are analytically little tractable in our framework. In particular, no closed-form solutions for the equilibrium preference parameters can be derived, except for the logarithmic benefit function for which $\theta_i^C = 1/3$. For the isoelastic form with $\eta_i \neq -1$, θ_i^C depends on β_i, as Figure 6(a) illustrates. In particular, θ_i^C approaches $1/3$ for $\beta_i \to 0$, and it approaches zero for $\beta_i \to 1$.[32]

[30] All computations and Mathematica files for this section are available upon request.

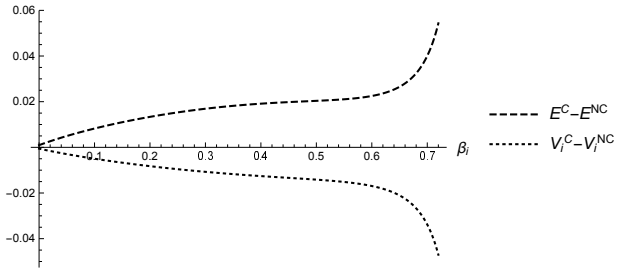
[31] The quadratic benefit function has an advantage over other specifications in terms of interpretation: no-policy emissions are those for which it holds: $B_i'(e_i) = 0$, i.e., no-policy emissions are finite. Therefore, it is also easier to specify what abatement is: the reduction of emissions relative to no-policy emissions. For the specifications we use in this section, $B_i'(e_i) \neq 0$ for all $e_i \geq 0$, and thus no-policy emissions go towards infinity.

[32] Notice that with a linear benefit function and a linear damage function as would be the case for $\beta_i = 1$, we have a degenerate solution: for the regime D, the first-order condition $B_i'(\omega_i) - \theta_i D_i' = 0$ (or $\alpha_i = \theta_i \delta_i$) either holds with equality for all levels of ω_i or it does not hold for any ω_i. This is different from Buchholz et al. (2015) where a convex damage function is assumed.

Figure 6: Symmetric equilibrium for isoelastic benefit functions (with $\alpha_i = 1/10$ and $\delta_i = 1$)



(a) Preference parameters



(b) Difference in aggregate emissions and welfare btw. regimes C and NC

Thus, the equilibrium preference parameters imply that the strategic delegation incentives are stronger for the isoelastic specification as compared to the quadratic benefit function, for which we obtained $\theta_i^C = 4/7$. This leads to *higher* aggregate emissions and *lower* individual welfare of the principals in regime C as compared to the quadratic specification as well as compared to the regime NC, as indicated by Figure 6(b).[33] All in all, we find that, at least for symmetry, isoelastic specifications of the benefit function deteriorate the prospects of mutually beneficial cooperation in the form of an international permit market. Strategic delegation incentives are particularly severe for high levels of β_i.

8.2 Polynomial benefit functions

Next, we examine more general polynomial functions of the form

$$B_i(e_i) = \frac{e_i}{\phi_i}(\epsilon_i - \frac{1}{2}e_i^{\beta_i}) \quad \text{with} \quad \beta_i > 0 \ . \tag{24}$$

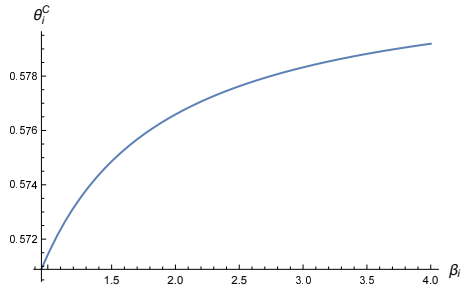
For the quadratic function employed earlier in the paper, $\beta_i = 1$. Note that, as before, ϵ_i has to be sufficiently large such that $B_i'(e_i) > 0$.

Figure 7 depicts the equilibrium preference parameter of the selected agent in regime C and the differences in equilibrium emissions and equilibrium welfare of the principal

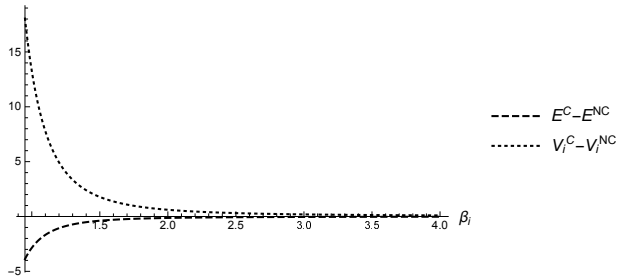
[33] For the logarithmic specification, we obtain closed-form solutions: $E^C - E^{NC} = \alpha_i/\delta_i > 0$ and $V_i^C - V_i^{NC} = \alpha_i(\ln(3/2) - 1) < 0$.

between the regimes C and NC.[34] We observe from Figure 7(a) that θ_i^C slightly increases with β_i. Moreover, in line with our earlier results for symmetric countries, aggregate emissions are always lower and individual welfare of the principals is always higher (Figure 7b) under cooperation as compared to non-cooperation, even for high values of β_i.

Figure 7: Symmetric equilibrium for polynomial benefit functions (with $\delta_i = 10$, $\phi_i = 1$ and $\epsilon_i = 100$)



(a) Preference parameters



(b) Difference in aggregate emissions and welfare btw. regimes C and NC

This robustness check regarding the benefit function thus shows that the curvature of this function does indeed play an important role. First, the quadratic specification might not give the most pessimistic results in terms of whether cooperation can improve upon the non-cooperative outcome, since under an isoelastic benefit function, the opposite is true. Second, for polynomial functions, our results turn out to be robust. Ultimately, it remains an empirical question what the correct specification of the benefit function is. While the widely used RICE model by Nordhaus (2010) assumes an almost cubic abatement cost function (with an exponent to abatement of 2.8), Cline (2011) finds that estimates of region-specific abatement cost curves based on data from leading "top down" models in the Stanford Energy Modeling Forum exhibit exponents of between 1.158 and 1.764 if assuming an abatement cost function of the form αa^β, where a is the level of abatement. Our quadratic specification lies in between these functional form assumptions respectively estimates.

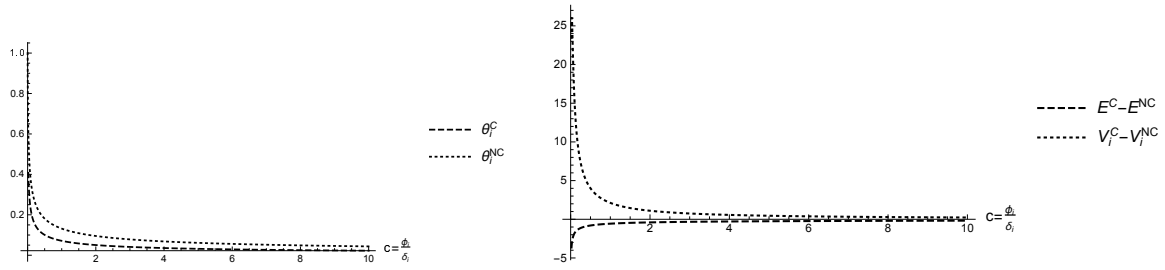
[34] We start at a value of $\beta_i = 0.95$ due to problems with scale (note that emissions soar towards infinity as $\beta_i \to 0$ in either regime, and in the regime NC faster than for C).

8.3 Quadratic damage function

As mentioned earlier, we employed a linear damage function for two reasons. First, this specification can be regarded as a reasonable approximation to reality, particularly in the short to medium run. Second, it eliminates an additional source of strategic interaction in the model and thus poses the least favorable conditions for strategic delegation to occur. In other papers, the assumed convexity of the damage function is the only source of strategic interaction (e.g., Siqueira, 2003; Buchholz et al., 2005), and it is therefore interesting to show whether a quadratic damage function changes the main results in our setting. To this end, we employ the following damage function: $D_i(E) = \delta_i E^2$, which is also often used in the literature on international environmental agreements. As before, we focus on the symmetric case, as closed-form solutions do not exist for all regimes.

Intuitively, the most obvious change is that permit choices by the agents in the threat point of the negotiations and in the non-cooperative regime are now strategic substitutes. In the non-cooperative regime, principals thus do not choose self-representation anymore. Furthermore, one would expect that the strategic substitutability in the choice of the agents' preference parameters by the principals becomes "stronger", as the principals have an additional reason to shift the burden of abatement to the other country (their marginal damage is not constant anymore).

Figure 8: Comparisons in symmetric equilibrium for quadratic damages (with $\delta_i = 5$ and $\epsilon_i = 5$)



(a) Comparison of preference parameters in regimes C and NC

(b) Difference in aggregate emissions and welfare btw. regimes C and NC

Figure 8 depicts the equilibrium preference parameters, emissions and welfare for the regimes C and NC for different ratios of $c = \phi_i/\delta_i$ (as before, the choice of ϵ_i has a level effect but does not affect the regime comparison). We observe from Figure 8(a) that, unlike for the case with linear damages, the equilibrium preference parameter in

the symmetric equilibrium varies with c in a straightforward way. The higher is δ_i as compared to ϕ_i, the "greener" is the agent selected by the principal in either regime. Furthermore, the agents chosen by the principals under cooperation exhibit strictly lower environmental preferences than the agents selected under non-cooperation. This, however, does not imply that aggregate emissions are higher under cooperation than under non-cooperation, as Figure 8(b) demonstrates. The reason for this is that emission levels are chosen in different ways (negotiations vs. non-cooperative Nash). Most importantly, due to lower aggregate emissions, cooperation results in *higher* individual welfare in the symmetric equilibrium than non-cooperation (Figure 8b), unless for values of $c < 0.02$ for which high emissions together with a low ϕ_i lead to very high positive welfare (not depicted). Thus, our findings obtained for linear damages (lower aggregate emissions and higher individual welfare under cooperation) also carry through for the case of quadratic damages, at least for symmetry. This does not imply, of course, that the same is true for *all* strictly convex damage functions, in particular higher-order polynomials or other functional forms. The shape of the benefit function does, however, seem to have a larger influence on the strength of strategic delegation and thus on whether cooperation is beneficial.

9 Conclusion

Given the very few international carbon markets (there exist only the EU emissions trading scheme linking all EU countries plus Iceland, Liechtenstein and Norway, and an international market linking California and Québec[35]), one may wonder why countries are so reluctant to establish international carbon markets. In this paper, we identify strategic politico-economic reasons in negotiations and transfer payments induced by the international carbon market as a potential explanation for the observed reluctance towards linking.

We find that the principals of two negotiating countries have an incentive to appoint agents that care less about environmental damages than they do themselves. Even with a linear damage function, the potential gains from permit trading in the international market are sufficient to create strategic considerations in the delegation

[35] Ontario had also participated in the latter market but revoked its commitment after a very short period of time. Furthermore, the EU and Switzerland signed an agreement in 2017 to link their systems starting from 2020.

decision. The good news is that strategic delegation does not fully erode the benefits from cooperation, at least for benefit and damage functions typically used in the literature on international environmental agreements. In particular, global emissions are still lower than if the principals were to choose policies non-cooperatively. We then ask whether countries benefit from cooperation in terms of welfare. Our results indicate that the countries' characteristics play a fundamental role in the success of international negotiations. When countries are almost identical in terms of marginal damages and marginal abatement costs or when they exhibit very different marginal damages, both principals gain from cooperation. For less extreme scenarios, we find that whether both countries benefit depends both on marginal damages and marginal abatement costs.

Moreover, we show that, when non-tradable instead of tradable emissions caps are negotiated, global emissions are lower and both principals are better off as compared to non-cooperation when the asymmetries are not too strong with respect to both marginal damages (as perceived by the principals) *and* marginal abatement costs. While a cooperative regime with non-tradable caps is a Pareto improvement over an international permit market only for few parameter combinations, non-tradable caps are more likely to lead to mutually beneficial agreements that are able to achieve emissions reductions. Our results thus suggest that transfers between countries – implicit in our model through the permit allocation – are not necessarily beneficial for successful negotiations as has often been suggested by the literature. They also caution against arguing for establishing more international permit markets on the grounds of economic efficiency. With regards to political economy, such markets might not last for long when governments realize that these markets are not necessarily in their best interest. We also showed that vastly diverging marginal abatement costs, which yield high efficiency gains from trading permits, are not sufficient for an international carbon market to be mutually beneficial. Finally, we showed that different functional form assumptions do play a role in our analysis. Although, for most specifications, our main messages remain unchanged in a symmetric setup, this exploration suggests that an interesting direction for future research would be to check the robustness of our results for different functional form specifications in an asymmetric setup.

Acknowledgments

We thank the editor Andreas Lange and two anonymous reviewers for extremely valuable comments and suggestions. We are also grateful to Nico Bauer, Antoine Loeper, Grischa Perino, Rick van der Ploeg, Ulrich Wagner, Ralph Winkler, seminar participants at the Universities of Hamburg, Umeå, and Ioannina, and conference participants at the APET 2017, FSR Climate Conference 2017, WCERE 2018, SURED 2018, CRETE 2018, ASSET 2019 and the PEGASOS workshop 2020 at the Potsdam Institute for Climate Impact Research.

Appendix

For ease of navigation this appendix is divided into sections that refer to the respective sections in the main text.

References to Section 4

Closed-form solutions at stage two:

Given our assumption about the functional forms, we find for the bargaining default:

$$\omega_i^D(\theta_i) = e_i^D(\theta_i) = \epsilon_i - \theta_i\delta_i\phi_i \ , \quad E^D(\Theta) = \epsilon - \theta_i\delta_i\phi_i - \theta_{-i}\delta_{-i}\phi_{-i} \ , \tag{A.1}$$

$$W_i^D(\Theta) = \frac{\epsilon_i^2}{2\phi_i} - \theta_i\delta_i(\epsilon - \frac{1}{2}\theta_i\delta_i\phi_i - \theta_{-i}\delta_{-i}\phi_{-i}) \ , \tag{A.2}$$

and for the NBS (making use of equation 7):

$$e_i^C(\Theta) = \epsilon_i - \phi_i(\theta_i\delta_i + \theta_{-i}\delta_{-i}) \ , \quad E^C(\Theta) = \epsilon - \phi(\theta_i\delta_i + \theta_{-i}\delta_{-i}) \ , \tag{A.3}$$

$$\omega_i^C(\Theta) = \epsilon_i - \frac{(\theta_i\delta_i)^2(4\phi_i + 3\phi_{-i}) + (\theta_{-i}\delta_{-i})^2\phi_i + 8\theta_i\delta_i\theta_{-i}\delta_{-i}\phi_i}{4(\theta_i\delta_i + \theta_{-i}\delta_{-i})} \ , \tag{A.4}$$

$$p^C(\Theta) = \theta_i\delta_i + \theta_{-i}\delta_{-i} \ , \quad T_i^C(\Theta) = \frac{3}{4}\Big((\theta_{-i}\delta_{-i})^2\phi_i - (\theta_i\delta_i)^2\phi_{-i}\Big) \ . \tag{A.5}$$

Differentiating the above solutions for both countries with respect to θ_i yields:

$$\frac{de_i^C(\Theta)}{d\theta_i} = -\delta_i\phi_i < 0 \ , \quad \frac{de_{-i}^C(\Theta)}{d\theta_i} = -\delta_i\phi_{-i} < 0 \ , \tag{A.6a}$$

$$\frac{dE^C(\Theta)}{d\theta_i} = -\delta_i\phi < 0 \ , \quad \frac{dp^C(\Theta)}{d\theta_i} = \delta_i > 0 \ , \quad \frac{dT_i^C(\Theta)}{d\theta_i} = -\frac{3}{2}\theta_i\delta_i^2\phi_{-i} \leq 0 \ , \tag{A.6b}$$

$$\frac{d\omega_i^C(\Theta)}{d\theta_i} = -\frac{\delta_i\left[\Big((\theta_i\delta_i)^2 + 2\theta_i\delta_i\theta_{-i}\delta_{-i}\Big)(4\phi_i + 3\phi_{-i}) + 7(\theta_{-i}\delta_{-i})^2\phi_i\right]}{4(\theta_i\delta_i + \theta_{-i}\delta_{-i})^2} < 0 \ , \tag{A.6c}$$

$$\frac{d\omega_{-i}^C(\Theta)}{d\theta_i} = -\frac{\delta_i\left[(\theta_i\delta_i + \theta_{-i}\delta - i)^2\phi_{-i} + 3(\theta_{-i}\delta_{-i})^2(\phi_{-i} - \phi_i)\right]}{4(\theta_i\delta_i + \theta_{-i}\delta_{-i})^2} \gtreqless 0 \ , \tag{A.6d}$$

$$\frac{d\left[\omega_i^C(\Theta) - e_i^C(\Theta)\right]}{d\theta_i} = -\frac{3\delta_i\left[(\theta_i\delta_i)^2\phi_{-i} + 2\theta_i\delta_i\theta_{-i}\delta_{-i}\phi_{-i} + (\theta_{-i}\delta_{-i})^2\phi_i\right]}{4(\theta_i\delta_i + \theta_{-i}\delta_{-i})^2} < 0 \ . \tag{A.6e}$$

As expected, total emissions as well as country-specific emissions increase when the selected agent in the considered country is less green, i.e., when θ_i is smaller. At the same time, the permit price unambiguously falls. The number of emission permits allocated to country i increases by even more than emissions do, so that country i is now more likely (not in a stochastic sense) to be the permit seller and less likely to be the permit buyer. Moreover, country i receives a (weakly) higher transfer or pays a (weakly) lower transfer to country $-i$ when θ_i is smaller.

References to Section 5

Proof of Proposition 3

The first part of the proposition regarding the comparison of the preference parameters of the selected agent and the principal ($\theta_i^C < 1 \quad \forall i = 1, 2$) can easily be seen, as the intercept of each reaction function is smaller than unity, and the slope of the reaction functions is negative.

We show the existence of a Nash equilibrium using Brouwer's Fixed Point Theorem:

The maximization problem of country i's principal is strictly concave:

$$\begin{aligned}\frac{d^2V_i^C(\Theta)}{d\theta_i^2} =& p^C(\Theta)\frac{d^2\omega_i^C(\Theta)}{d\theta_i^2} + \frac{dp^C(\Theta)}{d\theta_i}\left[2\frac{d\omega_i^C(\Theta)}{d\theta_i} - \frac{de_i^C(\Theta)}{d\theta_i}\right] \\ =& -\frac{1}{2}\delta_i^2(2\phi_i + 3\phi_{-i}) < 0 \ . \end{aligned} \tag{A.7}$$

This implies that this program has a unique maximum (if it has any). Moreover, from the Maximum Theorem it follows that the unique best response function of each principal is continuous in the strategy of the other principal, and as it is also compact (by the property that the set of preference parameters $[0, \theta_i^{\max}]$ is compact, i.e, bounded and closed), all conditions of the Brouwer's Fixed Point Theorem for the existence of a Nash Equilibrium are satisfied.

As both reaction functions are linear, we can theoretically have the following four cases, as illustrated by Figure 9. We define $\theta_i(\theta_{-i}^0) = 0$ and $\theta_{-i}(\theta_i^0) = 0$.

a) Unique interior Nash equilibrium if and only if:

$$\theta_i(0) < \theta_i^0 \quad \wedge \quad \theta_{-i}(0) < \theta_{-i}^0 \ . \tag{A.8}$$

Both conditions hold simultaneously if and only if:

$$\frac{2\phi_i}{3\phi_i + 2\phi_{-i}} < \frac{\delta_i}{\delta_{-i}} < \frac{2\phi_i + 3\phi_{-i}}{2\phi_{-i}} \ . \tag{A.9}$$

Note that the reaction functions also intersect exactly once, which is why there cannot be a continuum of interior Nash equilibria (along with two corner Nash equilibria). Plugging one reaction function into the other yields equation (15). This proves the first part of Proposition 3.

b) One interior Nash equilibrium and two corner Nash equilibria if and only if:

$$\theta_i(0) > \theta_i^0 \quad \wedge \quad \theta_{-i}(0) > \theta_{-i}^0 \ . \tag{A.10}$$

It can easily be shown that both conditions cannot hold simultaneously. Thus, no such equilibria exist.

c) Unique corner Nash equilibrium if and only if:

$$\theta_i^0 \leq \theta_i(0) \quad \wedge \quad \theta_{-i}(0) < \theta_{-i}^0 \ . \tag{A.11}$$

Both conditions hold simultaneously if and only if:

$$\frac{\delta_i}{\delta_{-i}} \geq \frac{2\phi_i + 3\phi_{-i}}{2\phi_{-i}} > 1 \ . \tag{A.12}$$

In this case, the equilibrium is described by $(\theta_i^C, \theta_{-i}^C) = \Big(2\phi/(2\phi + \phi_{-i}), 0\Big)$, where principal i exhibits the higher marginal damage, i.e., $\delta_i > \delta_{-i}$.

d) Unique corner Nash equilibrium if and only if:

$$\theta_i(0) < \theta_i^0 \quad \wedge \quad \theta_{-i}(0) \leq \theta_{-i}^0 \ . \tag{A.13}$$

Both conditions hold simultaneously if and only if:

$$\frac{\delta_i}{\delta_{-i}} \leq \frac{2\phi_i}{3\phi_i + 2\phi_{-i}} < 1 \ . \tag{A.14}$$

In this case, the equilibrium is described by $(\theta_i^C, \theta_{-i}^C) = \Big(0, 2\phi/(2\phi+\phi_i)\Big)$, where principal i exhibits the lower marginal damage, i.e., $\delta_i < \delta_{-i}$. Cases (c) and (d) thus prove the second part of Proposition 3.

□

References to Section 6

Delegation under non-cooperation among agents

If agents do not bargain but decide non-cooperatively about permit issuance on domestic permit markets (which they also do in the threat point of the negotiations), the principal in country i will select an agent with preference parameter θ_i in order to maximize

$$V_i^{NC} = B_i\Big(\omega_i^{NC}(\Theta)\Big) - D_i\Big(E^{NC}(\Theta)\Big) \ , \tag{A.15}$$

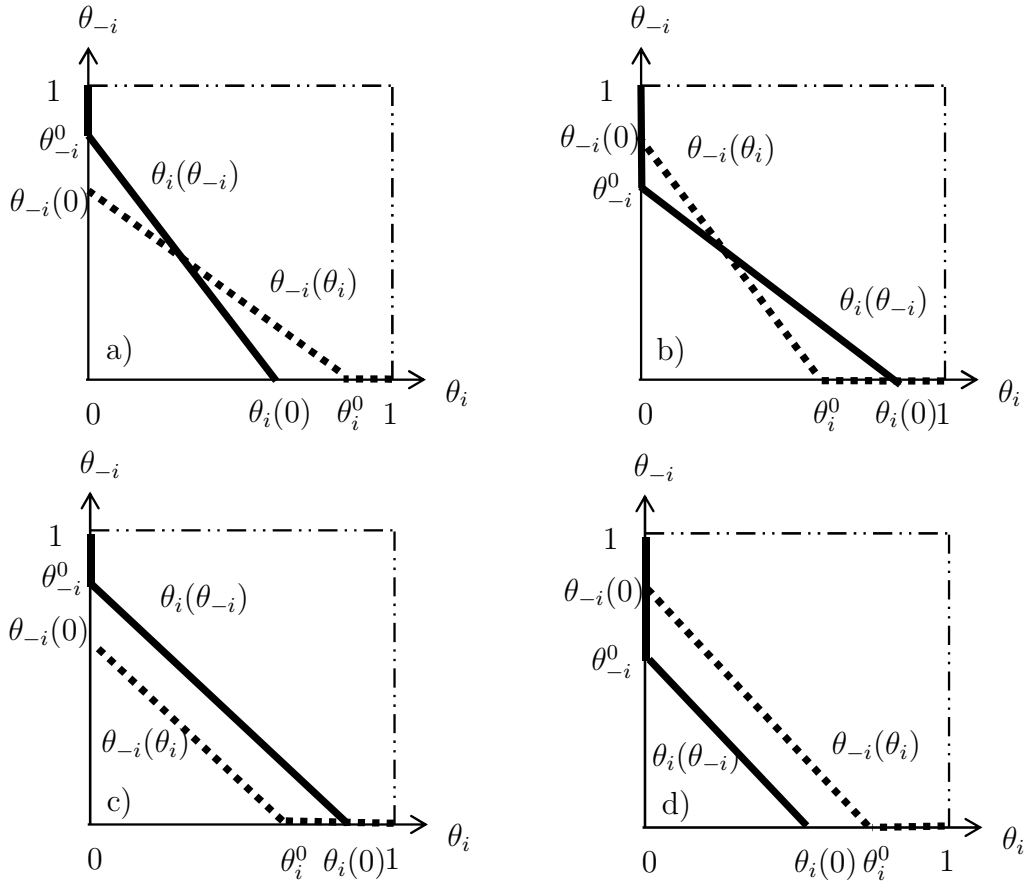


Figure 9: Potential Nash equilibria of the delegation stage.

given the Nash equilibrium $\Omega^{NC}(\Theta)$ of the subgame starting in the second stage as described by equations (8) and Proposition 1, and given the preference parameter θ_{-i} of the selected agent in the other country. The first-order condition gives us

$$B_i'(\omega_i^{NC}(\Theta))\frac{d\omega_i^{NC}(\Theta)}{d\theta_i} - D_i'\Big(E^{NC}(\Theta)\Big)\frac{dE^{NC}(\Theta)}{d\theta_i} = 0 \ , \tag{A.16}$$

which implicitly determines the reaction function of the principal in country i, $\theta_i^{NC}(\theta_{-i})$. Taking into account the equilibrium outcome in the second stage and in particular equation (8), the first-order condition becomes

$$(1-\theta_i)D_i'\Big(E^{NC}(\Theta)\Big)\frac{dE^{NC}(\Theta)}{d\theta_i} = 0 \ , \tag{A.17}$$

which implies that there is no incentive for strategic delegation: principals choose agents with the same preferences as theirs. We summarize this finding in the following proposition.

Proposition 6 (Unique Nash equilibrium at stage one under dom. markets) *When agents choose permit issuance on domestic permit markets in a non-cooperative way, there exists a unique Nash equilibrium at stage one,* $\Theta^{NC} = (\theta_i^{NC}, \theta_{-i}^{NC}) = (1,1)$ *in which the principals in both countries simultaneously choose agents with the same preferences as theirs: self-representation is the equilibrium strategy.*

Proof: As the reaction functions of the principals are orthogonal, there is exactly one point of intersection.

Substituting for $\theta_i^{NC} = \theta_{-i}^{NC} = 1$ in equations (A.1), we find:

$$\omega_i^{NC}(\theta_i) = \epsilon_i - \delta_i\phi_i \ , E^{NC} = \epsilon - \delta_i\phi_i - \delta_{-i}\phi_{-i} \ . \tag{A.18}$$

It thus holds that $\omega_i^{NC} = \omega_i^D(\Theta^P)$ and $E^{NC} = E^D(\Theta^P)$.

Proof of Proposition 4

To show that global emissions are lower in the equilibrium under delegated cooperation, $E^C(\Theta^C)$, as compared to the purely non-cooperative equilibrium, $E^D(\Theta^P)$, in which principals choose emissions permits for their domestic markets non-cooperatively, we have to distinguish two cases. In both cases, we plug $\theta_i^P = \theta_{-i}^P = 1$ into equation

(A.1) and obtain:

$$E^D(\Theta^P) = \epsilon - (\delta_i\phi_i + \delta_{-i}\phi_{-i}) \ . \tag{A.19}$$

1. Unique interior NE under delegated cooperation:
   In an interior equilibrium, we insert equation (15) for each $i = 1, 2$ into equation (A.3) and obtain:

$$E^C_{\text{interior}}(\Theta^C) = \epsilon - \frac{2\phi^2(\delta_i\phi_i + \delta_{-i}\phi_{-i})}{2\phi_i^2 + 2\phi_{-i}^2 + 3\phi_i\phi_{-i}} < E^D(\Theta^P) \ . \tag{A.20}$$

2. Unique corner NE under delegated cooperation:
   Assume that $\delta_i > \delta_{-i}$ such that $\theta_i^C = 2\phi/(2\phi + \phi_{-i})$ and $\theta_{-i}^C = 0$. Plugging these equilibrium preference parameters of the delegated agents into equation (A.3) yields:

$$E^C_{\text{corner}}(\Theta^C) = \epsilon - \frac{2\delta_i\phi^2}{2\phi + \phi_{-i}} < E^D(\Theta^P) \ . \tag{A.21}$$

□

Proof of Proposition 5

In what follows, we show for each of the cases listed in this proposition that these are mutually beneficial for the principals:

1. In this case, $a = \delta_i/\delta_{-i} = 1$. By Proposition 3, we thus have a unique interior Nash equilibrium at the first stage. Plugging the equilibrium preference parameters of the agents into the equation governing the permit price and the respective equations for emissions and allowance choices under both regimes C and NC, the payoff difference for principal i between these two regimes is given by:

$$\Delta V_i = \frac{\tilde{\delta}^2\left[4(\phi_i^2 + \phi_{-i}^2) + 5\phi_i\phi_{-i}\right]\left[2(\phi_i^3 - \phi_{-i}^3) + 3\phi_i^2\phi_{-i}\right]}{6\left[2(\phi_i^2 + \phi_{-i}^2) + 3\phi_i\phi_{-i}\right]^2} \ , \tag{A.22}$$

   where $\delta_i = \delta_{-i} \equiv \tilde{\delta}$.

The above payoff difference can be positive or negative. In particular, we have:

$$\text{for} \quad \phi_i \geq \phi_{-i}: \quad \Delta V_i > 0 \quad \wedge \quad \Delta V_{-i} \lesseqgtr 0 \ , \tag{A.23a}$$

$$\text{for} \quad \phi_i < \phi_{-i}: \quad \Delta V_i \lesseqgtr 0 \quad \wedge \quad \Delta V_{-i} > 0 \ . \tag{A.23b}$$

That is, for the two intervals, the principal of one country always benefits from cooperation, while the other one may or may not be better off under this regime compared to the regime in which policies are chosen in a purely non-cooperative way. It is straightforward to show that both principals will only be better off under cooperation when ϕ_i and ϕ_{-i} are sufficiently similar. For $\phi_i > \phi_{-i}$, it additionally must hold $\phi_i^3 - \phi_{-i}^3 < 3/2\phi_i\phi_{-i}^2$, while for $\phi_i < \phi_{-i}$, it must hold $\phi_{-i}^3 - \phi_i^3 < 3/2\phi_i^2\phi_{-i}$ (see last factor in the numerator of equation A.22).

2. In this case, $b = \phi_i/\phi_{-i} = 1$. By Proposition 3, we can either have a unique interior Nash equilibrium or a unique corner Nash equilibrium at the first stage. Defining $\phi_i = \phi_{-i} \equiv \tilde{\phi}$, the payoff differences for the principals in country i and $-i$ are given by:

$$\begin{aligned}
&(\Delta V_i, \quad \Delta V_{-i}) = \\
&= \begin{cases}
\left(\frac{\tilde{\phi}}{50}(8\delta_{-i}^2 + 30\delta_i\delta_{-i} - 25\delta_i^2), \frac{\delta_{-i}\tilde{\phi}}{10}(3\delta_{-i} - 10\delta_i)\right) & \text{for} \quad a \leq \frac{2}{5} \ , \\
\left(\frac{\tilde{\phi}}{98}(4\delta_{-i} - 3\delta_i)(3\delta_i + 10\delta_{-i}), \frac{\tilde{\phi}}{98}(4\delta_i - 3\delta_{-i})(3\delta_{-i} + 10\delta_i)\right) & \text{for} \quad \frac{2}{5} < a < \frac{5}{2} \ , \\
\left(\frac{\delta_i\tilde{\phi}}{10}(3\delta_i - 10\delta_{-i}), \frac{\tilde{\phi}}{50}(8\delta_i^2 + 30\delta_{-i}\delta_i - 25\delta_{-i}^2)\right) & \text{for} \quad a \geq \frac{5}{2} \ .
\end{cases}
\end{aligned} \tag{A.24}$$

It can easily be shown that ΔV_i and ΔV_{-i} are both strictly larger than zero in a corner equilibrium for $a < 3/10$ (or for $a > 10/3$ in the other corner equilibrium), while in the case of an interior equilibrium both principals benefit for $3/4 < a < 4/3$. Note that for $a = b = 1$, we have $\Delta V_i = \frac{13}{98}\tilde{\delta}^2\tilde{\phi} > 0$, which implies that both principals always benefit from cooperation in the symmetric setup.

3. In this case, $a \to 0$, i.e., we have, by Proposition 3, a unique corner equilibrium at the first stage. Taking the limit of the principals' payoffs for $a \to 0$, we get:

$$\lim_{a\to 0} \Delta V_i = \frac{b(1+b)^2}{(2+3b)^2}\delta_{-i}^2\phi_{-i} > 0 \ , \tag{A.25a}$$

$$\lim_{a\to 0}\Delta V_{-i} = \frac{b(1+2b)}{2(2+3b)}\delta_{-i}^2\phi_{-i} > 0\ . \tag{A.25b}$$

Thus, both principals benefit from cooperation.

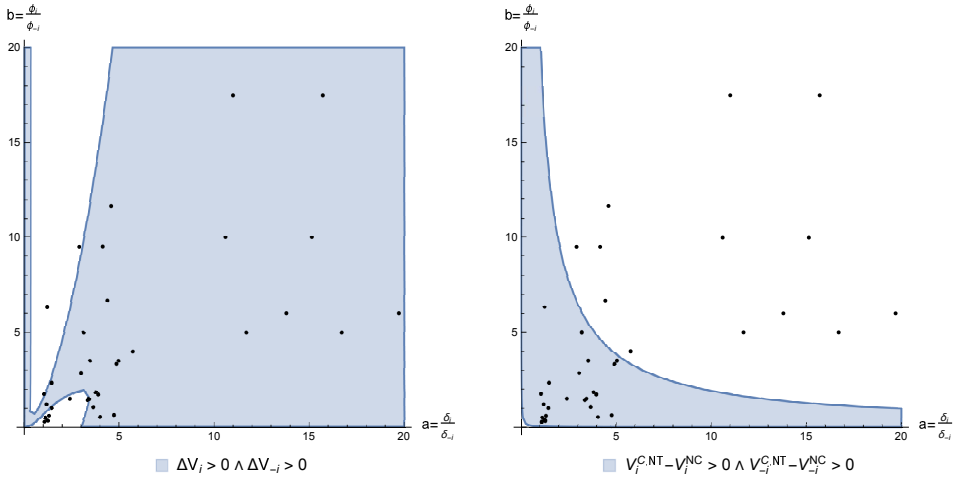
4. In this case, $b \to 0$, i.e., whether we have a unique corner or unique interior equilibrium now depends on a. Taking the limit of the principals' payoffs for $b \to 0$, we obtain:

$$\begin{aligned}&(\lim_{b\to 0}\Delta V_i,\quad \lim_{b\to 0}\Delta V_{-i}) = \\ &= \begin{cases} \left(-\frac{1}{3}a^2\delta_{-i}^2\phi_{-i}, \frac{1}{3}a^2\delta_{-i}^2\phi_{-i}\right) & \text{for} \quad 0 < a < \frac{3}{2}\ , \\ \left(\frac{1}{3}a(a-3)\delta_{-i}^2\phi_{-i}, \frac{1}{18}\delta_{-i}^2\phi_{-i}(2a^2+12a-9)\right) & \text{for} \quad a \geq \frac{3}{2}\ . \end{cases}\end{aligned} \tag{A.26}$$

Thus, for $b \to 0$, both principals benefit from cooperation if and only if $a > 3$ (and, of course, for $a \to 0$ as in case 3). □

Pairwise combinations of countries/country blocks

Figure 10: Parameter space for which cooperation is beneficial (the black dots represent all pairwise country/country block combinations from Table 2)



(a) Parameter space for which C is mutually beneficial

(b) Parameter space for which C, NT is mutually beneficial

References to Section 7

Stage two comparative statics for regime C, NT

Applying our functional form assumptions to equations (19), we arrive at the following results:

$$\omega_i^{C,NT}(\Theta) = \epsilon_i - \theta_i\delta_i\phi_i\left[1+\left(\frac{\theta_{-i}\delta_{-i}\phi_{-i}}{\theta_i\delta_i\phi_i}\right)^{\frac{1}{3}}\right] , \tag{A.27}$$

$$E^{C,NT}(\Theta) = \epsilon - \theta_i\delta_i\phi_i\left[1+\left(\frac{\theta_{-i}\delta_{-i}\phi_{-i}}{\theta_i\delta_i\phi_i}\right)^{\frac{1}{3}}\right] - \theta_{-i}\delta_{-i}\phi_{-i}\left[1+\left(\frac{\theta_i\delta_i\phi_i}{\theta_{-i}\delta_{-i}\phi_{-i}}\right)^{\frac{1}{3}}\right] . \tag{A.28}$$

Differentiate equations (A.27) and (A.28) with respect to θ_i:

$$\frac{d\omega_i^{C,NT}(\Theta)}{d\theta_i} = -\delta_i\phi_i\left[1+\frac{2}{3}\left(\frac{\theta_{-i}\delta_{-i}\phi_{-i}}{\theta_i\delta_i\phi_i}\right)^{\frac{1}{3}}\right] < 0 , \tag{A.29a}$$

$$\frac{d\omega_{-i}^{C,NT}(\Theta)}{d\theta_i} = -\frac{1}{3}\delta_i\phi_i\left(\frac{\theta_{-i}\delta_{-i}\phi_{-i}}{\theta_i\delta_i\phi_i}\right)^{\frac{2}{3}} \leq 0 , \tag{A.29b}$$

$$\frac{dE^{C,NT}(\Theta)}{d\theta_i} = -\delta_i\phi_i\left[1+\frac{2}{3}\left(\frac{\theta_{-i}\delta_{-i}\phi_{-i}}{\theta_i\delta_i\phi_i}\right)^{\frac{1}{3}}+\frac{1}{3}\left(\frac{\theta_{-i}\delta_{-i}\phi_{-i}}{\theta_i\delta_i\phi_i}\right)^{\frac{2}{3}}\right] < 0 . \tag{A.29c}$$

Like for regime C, if principal i chooses a marginally less green agent for the negotiations, the number of permits allocated to that country and allocated to the other country increase, resulting in an increase in aggregate emissions.

Existence of Nash equilibrium at stage one for regime C, NT

We show the existence of a Nash equilibrium at stage one in the regime without an international market, using Brouwer's Fixed Point Theorem.

The maximization problem of country i's principal is strictly concave:

$$\begin{aligned}\frac{d^2V_i^{C,NT}(\Theta)}{d\theta_i^2} =& B_i''\left(\frac{d\omega_i^{C,NT}}{d\theta_i}\right)^2 + B_i'\frac{d^2\omega_i^{C,NT}}{d\theta_i^2} - D_i'\frac{d^2E^{C,NT}}{d\theta_i^2} \\ =& B_i''\left(\frac{d\omega_i^{C,NT}}{d\theta_i}\right)^2 + D_i'\left[\frac{dE^{C,NT}/d\theta_i}{d\omega_i^{C,NT}/d\theta_i}\frac{d^2\omega_i^{C,NT}}{d\theta_i^2} - \frac{d^2E^{C,NT}}{d\theta_i^2}\right] < 0 ,\end{aligned} \tag{A.30}$$

where we made use of the first-order condition (21) and the comparative statics given in equations (A.29a) and (A.29c). This implies that this program has a unique maximum (if it has any). Moreover, from the Maximum Theorem it follows that the unique best response function of each principal is continuous in the strategy of the other principal, and as it is also compact (by the property that the set of preference parameters $[0, \theta_i^{\max}]$ is compact, i.e, bounded and closed), all conditions of the Brouwer's Fixed Point Theorem for the existence of a Nash Equilibrium are satisfied.

References

Anger, N. (2008). Emissions trading beyond Europe: Linking schemes in a post-Kyoto world. *Energy Economics 30*(4), 2028–2049.

Babiker, M., J. Reilly, and L. Viguier (2004). Is international emissions trading always beneficial? *The Energy Journal 25*(2), 33–56.

Buchholz, W., A. Haupt, and W. Peters (2005). International environmental agreements and strategic voting. *Scandinavian Journal of Economics 107*, 175–195.

Christiansen, N. (2013). Strategic delegation in a legislative bargaining model with pork and public goods. *Journal of Public Economics 97*, 217–229.

Cline, W. (2011). *Carbon Abatement Costs and Climate Change Finance*. Peterson Institute for International Economics.

Doda, B., S. Quemin, and L. Taschini (2019). Linking permit markets multilaterally. *Journal of Environmental Economics and Management 98*, 1–25.

Doda, B. and L. Taschini (2017). Carbon dating: When is it beneficial to link ETSs? *Journal of the Association of Environmental and Resource Economists 4*(3), 701–730.

Fershtman, C. and K. Judd (1987). Equilibrium incentives in oligopoly. *American Economic Review 77*, 927–40.

Flachsland, C., R. Marschinski, and O. Edenhofer (2009). To link or not to link: Benefits and disadvantages of linking cap-and-trade systems. *Climate Policy 9*, 358–372.

Gavard, C., N. Winchester, and S. Paltsev (2016). Limited trading of emissions

permits as a climate cooperation mechanism? US-China and EU-China examples. *Energy Economics 58*, 95–104.

Gerlagh, R. and M. Liski (2018). Consistent climate policies. *Journal of the European Economic Association 16*(1), 1–44.

Gersbach, H. and R. Winkler (2011). International emission permit markets with refunding. *European Economic Review 55*(6), 759–773.

Golosov, M., J. Hassler, P. Krusell, and A. Tsyvinski (2014). Optimal taxes on fossil fuel in general equilibrium. *Econometrica 82*(1), 41–88.

Graziosi, G. R. (2009). On the strategic use of representative democracy in international agreements. *Journal of Public Economic Theory 11*, 281–296.

Green, J. F., T. Sterner, and G. Wagner (2014). A balance of bottom-up and top-down in linking climate policies. *Nature Climate Change 4*, 1064–1067.

Habla, W. and R. Winkler (2013). Political influence on non-cooperative international climate policy. *Journal of Environmental Economics and Management 66*, 219–234.

Habla, W. and R. Winkler (2018). Strategic delegation and international permit markets: Why linking may fail. *Journal of Environmental Economics and Management 92*, 244–250.

Harstad, B. (2008). Do side payments help? Collective decisions and strategic delegation. *Journal of the European Economic Association 6*(2/3), 468–477.

Harstad, B. (2010). Strategic delegation and voting rules. *Journal of Public Economics 94*, 102–113.

Harstad, B. (2016). The dynamics of climate agreements. *Journal of the European Economic Association 14*(3), 719–752.

Hattori, K. (2010). Strategic voting for noncooperative environmental policies in open economies. *Environmental and Resource Economics 46*, 459–474.

Helm, C. (2003). International emissions trading with endogenous allowance choices. *Journal of Public Economics 87*, 2737–2747.

Holtsmark, B. and M. L. Weitzman (2020). On the effects of linking cap-and-trade systems for CO_2 emissions. *Environmental and Resource Economics 75*, 615–630.

Jaffe, J., M. Ranson, and R. N. Stavins (2009). Linking tradable permit systems:

A key element of emerging international climate policy architecture. *Ecology Law Quarterly 36*, 789–808.

Kempf, H. and S. Rossignol (2013). National politics and international agreements. *Journal of Public Economics 100*, 93–105.

Lange, A. and C. Schwirplies (2017). (Un)fair delegation: Exploring the strategic use of equity rules in international climate negotiations. *Environmental and Resource Economics 67*, 505–533.

Loeper, A. (2014). Public good provision and cooperation among representative democracies. *Available at SSRN 2435117*.

Loeper, A. (2017). Cross-border externalities and cooperation among representative democracies. *European Economic Review 91*, 180–208.

Marchiori, C., S. Dietz, and A. Tavoni (2017). Domestic politics and the formation of international environmental agreements. *Journal of Environmental Economics and Management 81*, 115–131.

Marschinski, R., C. Flachsland, and M. Jakob (2012). Sectoral linking of carbon markets: A trade-theory analysis. *Resource and Energy Economics 34*(4), 585–606.

McKinsey (2009). Pathways to a low-carbon economy. Version 2 of the global greenhouse gas abatement cost curve.

Montgomery, W. D. (1972). Markets in licenses and efficient pollution control programs. *Journal of Economic Theory 5*(3), 395–418.

Nordhaus, W. (2010). Economic aspects of global warming in a post-Copenhagen environment. *Proceedings of the National Academy of Sciences of the United States of America 107*(26), 11721–26.

Nordhaus, W. (2015). Climate clubs: Overcoming free-riding in international climate policy. *American Economic Review 105*(4), 1339–70.

Nordhaus, W. D. and J. Boyer (2000). *Warming the World: Economic Models of Global Warming Policies*. MIT Press.

Perino, G. (2010). How delegation improves commitment. *Economics Letters 106*, 137–139.

Persson, T. and G. Tabellini (1992). The politics of 1992: Fiscal policy and European

integration. *The Review of Economic Studies 59*, 689–701.

Roelfsema, H. (2007). Strategic delegation of environmental policy making. *Journal of Environmental Economics and Management 53*, 270–275.

Schelling, T. C. (1980). *The strategy of conflict*. Harvard University Press.

Segendorff, B. (1998). Delegation and threat in bargaining. *Games and Economic Behavior 23*, 266–283.

Siqueira, K. (2003). International externalities, strategic interaction, and domestic politics. *Journal of Environmental Economics and Management 45*, 674–691.

Sklivas, S. D. (1987). The strategic choice of managerial incentives. *The RAND Journal of Economics 18*(3), 452–458.

Spycher, S. and R. Winkler (2019). Strategic delegation in the formation of international environmental agreements. *Available at SSRN 3361731*.

Vickers, J. (1985). Delegation and the theory of the firm. *The Economic Journal 95*, 138–147.