Side-payments: An effective instrument for building climate clubs?

Håkon Sælen
CICERO Center for International Climate and Environmental Research, Oslo

Published in in *International Environmental Agreements: Politics, Law and Economics, 16* (6), pp. 909-932

Abstract: Climate clubs have been suggested as a gateway to substantial reductions in global emissions. The club approach begins with a small number of enthusiastic countries. This paper asks under what conditions such clubs are likely to evolve into effective cooperation through side-payments to new members. The question is addressed through a range of formal thought experiments using numerical simulations. The model is calibrated using empirical data on countries' emissions, GDP, populations, and vulnerabilities. It is simple and stylized, but allows for complex and dynamic interactions between actors. Basic equity considerations can be accommodated. The results indicate that side-payments' theoretical potential for facilitating effective clubs is large. One or two large emitters can initiate a club that grows to cover a substantial share of global emissions if the global benefit-cost ratio for mitigation is around 3 or larger. The size of stable clubs is larger if new members contribute to making side-payments, and somewhat lower if equity considerations constrain the set of possible transfers. Side-payments' effect is enabled by the large asymmetries between countries. Total side-payment flows range from tens to hundreds of billions of US dollars annually.

Key Words: climate change; climate clubs; international environmental agreements; side-payments; international transfers; agent-based modeling.

Acknowledgements: I would like to thank particularly Jon Hovi, Arild Underdal, and Detlef Sprinz, who have inspired me and helped me greatly in this work. Parts of this paper were presented to the 2014 ISA convention (panel WC30). I am indebted to Xinyuan Dai, Nils Weidmann, Torben Mideksa, the participants at the ISA panel, and two anonymous reviewers for helpful comments and suggestions. Finally, I thank Frank Azevedo for excellent copy-editing. Any remaining errors and omissions are my own. This work has been funded by the Norwegian Research Council through Strategic Challenges in International Climate and Energy Policy (CICEP).

1 Introduction

Multiple contemporary developments provide rationales for increasing scholarly attention to whether actions by individual countries and small groups thereof can grow into effective mitigation of global emissions. The first development is the shift away from internationally negotiated commitments to "nationally determined contributions" as the building block for the new global agreement. Second, that agreement – to be reached in Paris in December 2015 – will unlikely deliver the ambition many hope for. In particular, the parties' domestically determined contributions will not collectively be nearly ambitious enough to put us on a path to limiting warming to 2°C above pre-industrial levels – the goal Parties agreed to in Copenhagen in 2009[1] (UNFCCC 2009). This shortcoming makes it pertinent to look for processes outside the UNFCCC conference rooms to further boost ambition. Third, widely regarded as the most positive development in pre-Paris negotiations was the bilateral agreement between the United States and China announced on 12 November 2014 (see Goodell (2014) for a behind-the-scenes account). This event marks a new development in climate negotiations, which to date have been based on multilateral conference diplomacy. While not replacing the multilateral process under the UNFCCC, the bilateral initiative set commitments for the two largest emitters outside that process.

This paper seeks to identify conditions under which efforts started by a small group – or even a single actor – can effectively limit global emissions. More specifically, it investigates the conditions under which the initial group can use side-payments to attract additional actors to join their effort. The paper adds to the emerging literature on gradually evolving climate cooperation, which has been termed *the club approach* (Author et al. forthcoming).

[1] The Copenhagen Accord is not supported by all Parties, but the 2°C goal was universally agreed one year later as part of the Cancun Agreements.

A climate club based on side-payments can be described as a complex and dynamic phenomenon. Complexity arises from the large number of heterogeneous countries, whose behavior partly depends on the behavior of others. Among the world's countries, the numbers of possible clubs and bilateral side-payments are indeed very large. The dynamic aspect is important because club growth is crucial to effectiveness. These characteristics make it quite difficult to think through the implications of different assumptions without the aid of a formal model. They also imply difficulties for game-theoretic analyses. This paper therefore makes use of an agent-based model (ABM), a type of model particularly well suited for complex and dynamic phenomena.

The aim is to facilitate simple thought experiments, rather than to provide detailed predictions. For that purpose, it is desirable to keep the model's structure relatively simple. Too many model parameters can make it difficult to understand the interplay between them (Weidmann and Salehyan 2013). Into a relatively simple structure, agents resembling the world's countries are introduced, based on empirical data on GDP, emissions, populations, and vulnerabilities to climate change. Relative to the most geographically detailed game-theoretical model reviewed below (McGinty 2007), this setup increases the number of potential bilateral transactions more than 50 times. In addition, the current model includes an explicit process of club growth, whereas game-theoretic models only produce equilibrium outcomes.

The following section briefly outlines the extant literature on climate clubs and related ideas. Theoretical arguments for a club approach are discussed, and climate clubs' limited empirical success to date is noted. A research gap exists concerning the identification of the conditions under which clubs *would* be effective in reducing emissions. The subsequent section discusses insights on side-payments from the game-theoretic literature, which give some reason for optimism regarding their effectiveness when actors are heterogeneous. The paper then presents the ABM and its underlying data. A key feature is to assume that one or a small number of "enthusiastic" actors starts a club and attempts to recruit members through side-payments. The results and discussion section indicates that the largest emitters have considerable potential to initiate sizeable clubs, sometimes even single-handedly. This result is obtained for moderate assumptions about the global benefit-cost ratio of mitigation. Clubs' effectiveness are increased when new recruits contribute to funding additional side-payments, and reduced but not eliminated by constraints based on equity considerations. The total side-payments involved range from tens to hundreds of billions of US dollars annually. In the concluding remarks, side-payments are argued to compare favorably with other instruments for club growth in terms of theoretical effectiveness. However, side-payments face considerable obstacles in terms of political feasibility, and in ensuring they can be credibly made.

2 Climate clubs

Starting out with a small number of countries is the defining feature of climate clubs in this paper's conception of the term. In response to the developments in international climate diplomacy outlined in the introduction, the last five years have seen a marked increase in academic proposals for that form of regime. In a recent review of proposals, Falkner (2015) identifies three models with different rationales for how starting small may be beneficial: i) by making dialogue and bargaining simpler; ii) through creating incentives for membership; and iii) through increasing the legitimacy of the climate regime in the eyes of the great powers by giving them a privileged space. This paper focuses exclusively on the second rationale, and even more pointedly on side-payments as an incentive for membership.

The analysis draws in particular on a proposal by Victor (2011), who divides countries into two categories: enthusiastic and reluctant countries. Enthusiasts are willing to spend their own resources on climate mitigation and are the engine of international climate cooperation. Reluctant countries are those who will mitigate emissions only when such efforts coincide with other national goals. The key to broad and effective cooperation is that the enthusiastic countries must incentivize the reluctant countries to contribute. Inspired by the development of the GATT and the WTO, he proposes that a club initiated by enthusiasts deepens and expands through periodic bidding and negotiations with reluctant countries. Potential members would assemble an “accession” package of promises of what they would do to become a member, and existing members would offer incentives in return.

Victor discusses several forms of incentives. Carrots include money, technology cooperation, market access, security guarantees, and conditional additional mitigation efforts. Sticks, in the form of trade sanctions, are also considered. Victor does not advocate a carrot-only approach, arguing that it entails diplomatic outcomes that are impractical at the scale needed to make deep cuts in emissions. In particular, he cautions that Western nations will be wary of letting their government funds be used for cleaning up the energy systems in their foremost economic competitors, for example, Brazil and China. This paper will contribute to this analysis by quantifying both the effect monetary side-payments would likely have on climate cooperation, and the scale of the transfers needed. This quantification is done based on a formal model, where side-payments are required to benefit both donors and recipients. A full consideration of whether such side-payments are politically realistic is left to other works.

Also left to other works are other types of incentives for club membership. Another paper using the same model as the current paper (Author et al, forthcoming) analyzes conditional additional mitigation efforts and goods that benefit members exclusively, such as technological cooperation and preferential market access. It concludes that neither conditional mitigation commitments nor club-good benefits will likely effectively boost climate cooperation on their own; however, in combination they could substantially increase global mitigation under a relatively broad range of conditions. Negative incentives, in the form of trade penalties on participants, have been analyzed by Nordhaus (2015). Using game-theoretic simulations, he shows that a climate club[2] combining carbon pricing with import tariffs can produce high levels of participation and abatement. Contributions outside the club-specific literature have also shown that the threat of trade sanctions on non-participants can facilitate broad and stable cooperation to protect the environment under certain conditions (e.g., Barrett 1997, 1999; Helm et al. 2012; Neumayer 2001). Another potential basis for participation incentives is low-carbon technology. Stewart et al. (2013) argues that cooperation on research and development could produce benefits that intellectual property rights could ensure arise only to club members. A different mechanism is suggested by Urpelainen (2013a), who shows that a small number of countries may induce other countries to abate more in the future, through strategic investment in technology that lowers abatement costs.

The current paper describes a regime that grows stronger over time through gradually increasing membership. Other possibilities for regime strengthening are increasing the scope of sub-topics included or deepening participants' commitments. Falkner et al. (2010) proposes a "building blocks" approach that develops different elements of climate change governance in a step-wise fashion. Urpelainen (2013b) proposes a long-term strategy in which a series of "small wins" over time alter the payoffs of the international cooperation game to enable increasingly ambitious policies.

Empirical studies of minilateral efforts on climate change indicate that their achievements have been limited to fostering dialogue and discussions (Andresen 2014; Weischer et al. 2012). However, scholars have only just begun explaining *why* the efforts have failed in reducing emissions, and identifying the conditions under which climate clubs might grow and successfully address emissions. This paper contributes to filling that gap by employing a formal model inspired by Victor's club proposal to identify conditions for club effectiveness.

---

[2] Nordhaus' club diverges from this paper's definition of a club as starting with a small number of countries. Instead, club formation is a top-down process, where the regime is optimized to attract many participants and to achieve much abatement, and then countries decide whether to join.,

3 Game-theoretic models with side-payments

Distinct from the relatively recent literature on climate clubs, there is a well-established and extensive literature on the effect of side-payments, especially within game theory. Several game-theoretical models have analyzed the effect of side-payments – or *transfers*, as they are often termed – on international environmental agreements (IEAs). These models broadly but not unanimously conclude that side-payment possibilities lead to increased participation and abatement if actors are heterogeneous but not if they are homogeneous.

Two general classes of models dominate the game-theoretic literature on IEAs (Finus 2003). Terms used to identify the first class are *membership models*, *coalition models, two-stage games,* and *reduced-stage games.* The other class is termed *compliance models* or *repeated games*. Both strands of literature share the same basic framework, where the essence is that abatement leads to costs that are private but benefits that are shared by all in terms of reduced damage costs (a non-excludable and non-rival bad). This structure implies an incentive to free ride on others' abatement efforts, which leads to an (not necessarily unique) equilibrium with no cooperation. Most analyses of side-payments have been performed with membership models, so this review will focus on that strand. However, a limitation common to these models is the assumption that members comply with their obligations. In contrast, compliance is endogenous in compliance models. The second strand will therefore be brought in towards the end when discussing compliance.

3.1 Membership

Membership models include contributions from both cooperative game theory, which assumes binding agreements can be made without cost, and non-cooperative game theory, which does not make this assumption. A key contribution from cooperative game theory is by Chander and Tulkens (1995), who derive a simple side-payment scheme that sustains universal participation when countries are heterogeneous. Under this scheme, every country receives a side-payment that covers its increase in abatement costs from the non-cooperative equilibrium to the social optimum, and contributes to the total side-payments according to its share of benefits from emissions reduction (its marginal damage cost divided by the sum of all countries' marginal damage cost). The Chander–Tulkens rule has gained much attention, since it ensures that the agreement profits all countries even when they are heterogeneous. Most models in both cooperative and non-cooperative use a similar sharing rule. This can easily lead to politically unrealistic predictions. Because poorer countries are generally more vulnerable to climate change, they end up with a large share of the effort, sometimes even compensating the rich (e.g.,McGinty 2010). Where equitable schemes have been examined, they have been found less effective (Altamirano-Cabrera and Finus 2006). The ABM developed in this paper addresses equity concerns through an option that disallows transfers from a poorer to a richer country.

Cooperative game-theoretic models have been criticized in particular for making the strong assumption that countries believe that should they unilaterally withdraw, the whole coalition would dissolve (Barrett 2001; Finus 2003; McGinty 2007). This amounts to a severe punishment of non-participation, which may not be credible. These models also fail to explain the empirical pattern that most IEAs fall short of delivering efficiency (Finus 2003). The more recent literature is therefore dominated by non-cooperative game theory. Most contributions import the solution concept of internal and external stability from the study of cartel formation in oligopolies (d'Aspremont et al. 1983). This concept relies on weaker and more credible punishment of non-participation, since it does not assume complete coalition dissolution in response to individual withdrawal. A central general finding of this literature is what Nordhaus (2015) calls the "small coalition paradox": that coalitions tend to be either narrow and deep or broad and shallow. Furthermore, the stable coalition's size is a negative function of the potential gains from cooperation (Barrett 1994).

On side-payments, the main message is similar to that from Chander and Tulkens (1995): side-payments are effective among heterogeneous countries. A representative paper is by Barrett (2001), who finds that side-payments have virtually no effect when countries are homogeneous, but a dramatic positive effect when countries are strongly asymmetric. The model is very simple and stylized. There are only two types of countries – rich and poor – who make a binary decision – to pollute or abate. Abatement costs and benefits are linear, and are both assumed larger for rich than for poor countries. Under some parameter conditions, a stable coalition consists of all poor countries and a substantial share of rich countries. This finding has been questioned by Fujita (2006), who shows that the equilibrium proposed by Barrett exists only within a very narrow parameter range, and that his numerical example is in fact not an equilibrium. This finding perhaps illustrates Barrett's own statement that games of international cooperation between heterogeneous countries are notoriously difficult to work with.

Further illustrating the difficulty of developing analytical models, an analytical solution has only recently been found to the heterogeneity problem for two types of countries that have non-linear (quadratic) abatement costs (but linear damage costs) and choose strategies from a continuum, by Fuentes-Albero and Rubio (2010). They examine transfers separately in the context of heterogeneous abatement costs and heterogeneous damage costs. They find that side-payments have no effect if only abatement costs vary, but increase cooperation the more damage costs vary. Notably, this paper too has been accused of errors (Glanemann 2012).

Because of difficulties of solving analytical models of IEAs with heterogeneous actors, other studies rely (at least primarily) on numerical simulations. In a five-region model, Botteon and Carraro (1997) find that transfers among heterogeneous regions can increase welfare. With a 12-region model, Weikard et al. (2006) investigate eight sharing rules applied to the gains from cooperation, and Altamirano-Cabrera and Finus (2006) consider seven allocation rules applied to emission permits. Weikard et al. (2006) find small stable coalitions for all rules. Altamirano-Cabrera and Finus (2006) find stable coalitions for “pragmatic” rules (based on the current distribution of emissions, so-called grandfathering) but not for “equitable” rules. For comparison, Finus et al. (2006) find no stable coalitions without side-payments in the same model . In contrast, Nagashima et al. (2009), using an extended version of the same model, find that with grandfathered permits, the most efficient stable coalition performs worse than the most efficient coalition in the no transfers–scenario. Other transfer schemes do enhance effectiveness, but coalitions remain small and relatively ineffective. Even with optimal transfers, ambitious coalitions consisting of, for example, the United States, the EU15, and China are not stable.

Recent advances in the literature have also been made by McGinty, in terms of extending simulations to twenty regions (2007), alternative equilibrium concepts and game structures (2010), and new sharing rules (2011). However, the general conclusion persists that side-payments among heterogeneous countries can increase abatement and make all better off (2014).

3.2 Compliance

As noted, membership models do not explicitly address compliance with agreed obligations. This applies also to the model developed in this paper. With side-payments, there are two potential compliance problems, as Finus (2003) notes. The first is between donors and recipients. Either the recipient may be able to secure the payment without making the agreed effort, or it may make the agreed effort without receiving payment. The second is among donors, where there is an incentive to try to free ride on others’ donations. In a sense, side-payment possibilities transform one cooperation problem – that of abating – into another – one of contributing to side-payments.

Finus (2003) argues that side-payments reduce enforcement power over donors but increase it over recipients. Through non-compliance, a donor could benefit both by increasing its emissions and by avoiding side-payments. Therefore, a stronger deterrent against free riding is needed than in the no-side-payment case. Conversely, defecting recipients benefit by increasing emissions, but lose future side-payments. This asymmetry means that side-payments are an ideal instrument to credibly sanction developing countries.

4 Model description

The model aims to capture select essential features of climate clubs, while leaving out many real-world complicating factors. It is built on empirical data on countries' GDP, emissions, vulnerabilities, and populations. The basic decision is binary: Each actor must decide whether to be a member of a climate club or not.[3] Club obligations could be defined in many different ways. The model club is a specific conception of a climate club, but other conceptions would likely lead to very similar conclusions. It stipulates that all members undertake mitigation worth an equal percentage of their GDP. Side-payments compensate for parts of some actors' abatement costs, endogenizing the final cost distribution. The model allows equity-related restrictions to be placed on side-payments; for example, payments from a poorer to a richer country may be disallowed.

The model is a one-shot sequential game with an indefinite number of stages, where all costs are undiscounted and presented in per annum terms, and all actions that are not repealed are instantaneously implemented. While it cannot capture all real-life challenges to international cooperation on mitigating climate change, it provides a useful tool for systematically exploring the potential contributions to enhancing climate change mitigation by clubs based on side-payments..

4.1 Model procedures

[3] Because I model European Union members as a single unit, I use the term "actors" rather than "countries."

The Technical Appendix provides the model's formulae and describes the model in pseudo-code, thereby complementing the verbal description I offer here. The model and its input data can be downloaded via http://modelingcommons.org/browse/one_model/4458.

In short, the model explores scenarios in which small groups of different actors start a club and assess whether there are potential side-payment deals where the payment is less than the clubs' benefit from the entrant's mitigation but greater than the entrant's mitigation cost. Any such potential entrants join the club sequentially in return for a side-payment.

The model's actors are of two types, depending on their motivation for mitigation. *Reluctant* actors are rational and self-interested, and will join the club if and only if doing so leads to private benefits that exceed their abatement costs. *Enthusiastic* actors are defined by (Victor 2011) as willing to spend their own resources on mitigation. I assume they have an exogenous motivation to start a club and hence incur abatement costs even without any commitment by reluctant parties to follow suit. This assumption could be defended by arguing that state behavior is influenced by norms, values, and notions of collective identities that generate more cooperative behavior than one would expect based solely on actors' material self-interest (Mayer 1992; Underdal et al. 2012). However, the purpose for including such actors is not to describe the actual behavior of any real states, but to assess the likely effects of hypothetical leadership. I furthermore assume that enthusiastic actors will not necessarily abandon the club even if they would benefit by withdrawing unilaterally. However, I place one limitation on their enthusiasm: they will abandon the club if – after negotiating with all reluctant actors – the club generates negative net private benefits for that actor, *relative to the no-club scenario*. In this aspect, enthusiasts behave similarly to the *conditional cooperators* modeled by Richter and Grasman (2013) and to the *strong reciprocators* modeled by (Author 2014). Their preference structure conforms to that of an assurance game, in that they prefer to participate conditional on a certain amount of participation by others. This behavior is also similar to that assumed by cooperative game-theoretic models, where all members act as if the club would collapse if they left unilaterally. As discussed above, this assumption has been criticized by proponents of non-cooperative game theory, where actors consider whether the club would persist if they were to become a free rider. I assume the latter behavior only for reluctant actors: they will choose non-membership as long as their payoff outside the club (where they can free ride on members' mitigation) is greater than their payoff inside it. The reason I adopt the assumption from non-cooperative game theory for reluctant actors but the assumption from cooperative game theory for enthusiasts is that it would seem contrary to the very idea of enthusiasm to leave a club that would persist after the enthusiast had left.

Assuming actors measure outcomes solely in terms of impacts on GDP follows the interest-based approach to explaining international environmental regulation (Sprinz and Vaahtoranta 1994). The parsimony of this approach makes it particularly suitable for formal models. However, modeling states as unitary actors ignores that some of the greatest impediments to international cooperation are created by the interaction between domestic and international political processes (Mayer 1992). This weakness is shared with most formal models of international climate politics, with a rare exception provided by Kroll and Shogren (2008).

Actors have four further attributes with empirically grounded values: emissions (from fossil fuels and cement (Global Carbon Project 2014(2013 figures)), and from land-use change and forestry (World Resources Institute 2014 (2011 figures)), GDP (World Bank 2014a (2013 figures)), population (World Bank 2014b(2013 figures)), and climate-change vulnerability scores (Notre Dame Global Adaptation Index 2014 (2012 figures)). Complete data is available for 168 countries, accounting for 98% of both global emissions and gross global product (GGP). The European Union is modeled as a single actor; hence, the model includes 141 actors (or agents). Data for the twenty largest emitters are listed in Table 1.

Actor	Emissions (Share of global)	GDP (Share of GGP)	Vulnerability index
China	27.3	12.3	0.30
US	13.6	22.4	0.20
EU	9.0	23.2	0.20
India	6.4	2.5	0.43
Indonesia	4.8	1.2	0.34
Russian Federation	4.7	2.8	0.29
Japan	3.1	6.5	0.29
Brazil	2.2	3.0	0.30
Canada	1.8	2.4	0.23
Iran	1.7	0.5	0.29
South Korea	1.6	1.7	0.35
Saudi Arabia	1.5	1.0	0.34
Mexico	1.4	1.7	0.29
South Africa	1.3	0.5	0.37
Malaysia	1.1	0.4	0.31
Australia	1.0	2.1	0.24
Venezuela	0.9	0.6	0.29
Thailand	0.9	0.5	0.31
Kazakhstan	0.9	0.3	0.28
Turkey	0.8	1.1	0.28

Table 1: Emissions (including land-use change and forestry), GDP, and vulnerability index scores of the twenty largest emitters. Sources: Global Carbon Project (2014) (fossil fuel and cement emissions in 2013), World Resources Institute (2014) (land-use change and forestry emissions in 2011), World Bank (2014a) (GDP in 2013 at market exchange), and Notre Dame Global Adaptation Index (2014) (vulnerability scores for 2012). EU vulnerability is the average of its members' vulnerabilities.

4.1.2 Mitigation Costs and Benefits

Estimating the global costs and benefits of climate mitigation is beyond this study's scope. Instead, I run the model for different assumptions about club effectiveness in terms of damage costs avoided. The input variable *Global Damage* expresses the assumed benefit-cost ratio of a global club. I focus on a club where members undertake mitigation worth 1% of their GDP. I then test different assumptions about what benefits this club brings. It is possible to vary the cost too, for example by changing it to 2%, but what mostly matters for countries' motivation is the benefit-cost ratio, so I will keep costs fixed and vary benefits. The ratio's true value is inherently uncertain, and estimates also vary widely, as illustrated by the debate between Bill Nordhaus, Nicholas Stern, Martin Weitzmann, Richard Tol, and others. However, it turns out that in this model, the most interesting dynamics take place for ratios between 2 and 3, and that ratios outside this range typically lead to corner solutions. This means that a small number of model runs provides a good picture of what happens under all different assumptions about this specific input.

If only a subset of the actors participates, the total benefit generated by the club is assumed to be a linear function of the emissions covered. For example, a club covering 50% of global emissions is assumed to produce 50% of the climate benefits produced by a global club. This assumption rules out carbon leakage.

Finally, I assume that damage costs are distributed in proportion to actors' GDP and vulnerability. Because vulnerability affects damage costs, which in turn affect the incentive for club membership, vulnerability heterogeneity leads to heterogeneous incentives for membership. The model incorporates empirical data on vulnerability from the Notre Dame Global Adaptation Index (2014), which allocates scores based on actors' exposure, sensitivity, and adaptive capacity in eight sectors. The scores range from 0.15 (Switzerland) to 0.59 (Burundi) and are denoted $NDGAIN_i$. How to translate index scores into damage functions is a non-trivial question, for which limited empirical guidance is available. I address this challenge through a model input variable called *Vulnerability Weight,* which determines the variance of the damage-cost distribution across countries, while keeping constant global costs and the ranking of countries' damage costs (per unit of GDP). This constancy is achieved with the following formula:

$$Vulnerability_i = \frac{NDGAIN_i + (Vulnerability\ Weight - 1) \times (NDGAIN_i - \overline{NDGAIN})}{\overline{NDGAIN}} \qquad (1)$$

It expresses actor-specific vulnerability as the percentage loss in GDP_i arising when the global loss is 1% of GGP. $\overline{NDGAIN}$ is the GDP-weighted average NDGAIN score across all actors. Higher *Vulnerability*

Weight means higher variance in vulnerability across actors. I will run the model for the values 0, 1, and 2. When *Vulnerability Weight* equals zero, every actor's *Vulnerability* equals 1. In contrast, when *Vulnerability Weight* equals 1 (2), Switzerland's *Vulnerability* equals 0.6 (0.2) and Burundi's *Vulnerability* equals 2.3 (3.7). The distributions of vulnerability across actors for these parameters are displayed in Figures 1 and 2. To obtain actor-specific damage, I multiply $Vulnerability_i$ by *Global Damage*. Damage costs, as well as all other costs and benefits, are measured in percentage of GDP per annum. Given the above definitions and assumptions, an actor's incentive to mitigate climate change depends on its vulnerability and emissions. Emissions matter because (other things being equal) a large actor emits more GHGs than a smaller actor does and therefore causes more climate damage, including to itself.

The assumptions that an actor's membership costs are proportional to its GDP and the global benefits produced by its membership are proportional to its emissions imply that abatement costs per unit of emissions are proportional to GDP per unit of emissions. This simple relationship comes remarkably close to the relationship derived by Nordhaus (2015) through combining an integrated assessment model with estimates from an engineering model. However, because the club membership criterion is that spending on mitigation must equal 1% of GDP, variations in abatement costs do not affect incentives for club membership. This limited attention to abatement-cost variation is supported by Fuente-Albaro and Rubio's (2010) finding that this form of asymmetry has no important effect on the scope of cooperation.

Figure 1: Histogram of actors' damage costs (% of GDP) under baseline assumptions

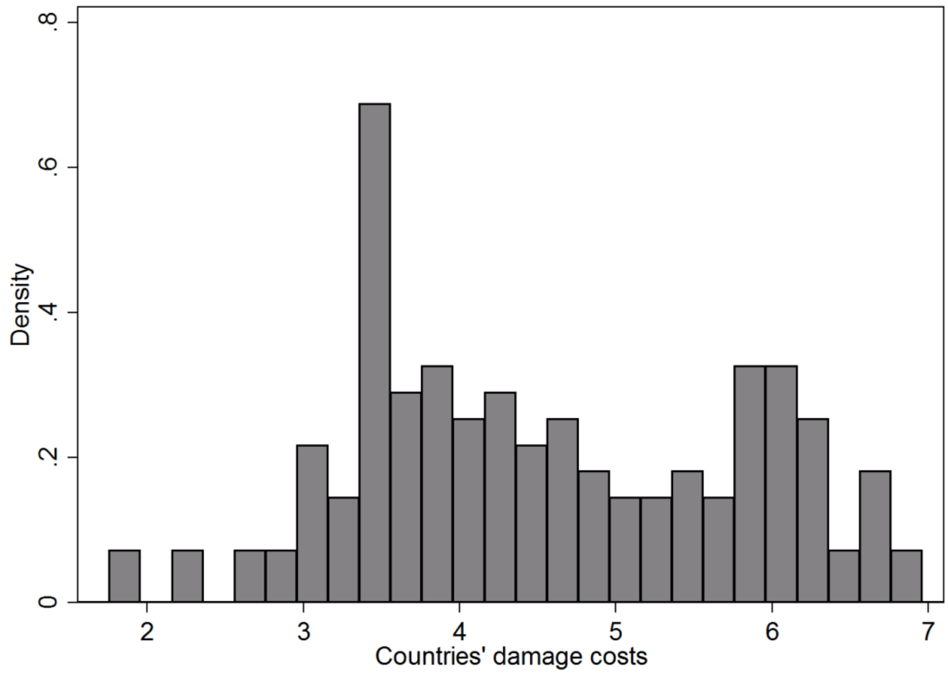
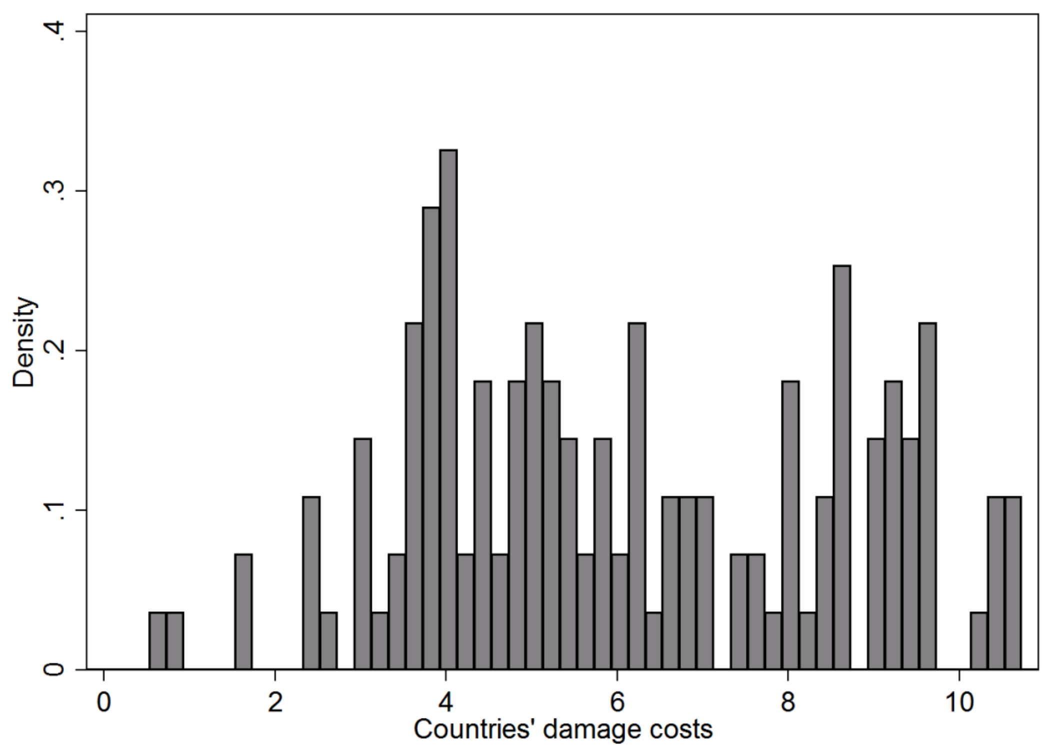


Figure 2: Histogram of actors' BAU damage costs (% of GDP) when Vulnerability Weight equals 2 and Global Damage Cost equals 3% of GGP



4.1.3 Side-payments

When a new member *i* enters the club, existing members benefit because the new member reduces emissions. If the sum of benefits to members would outweigh reluctant country *i's* net cost of joining (1% of GDP minus its share of benefits from its own mitigation), a mutually advantageous expansion is possible. I will assume *i* enters when such Pareto improvements are possible.

It should be noted that these assumptions ignore potential obstacles presented by bargaining over how to split the gains from cooperation between existing members, and between existing members and the club. Existing members face an assurance game with multiple equilibria where either all pay zero, or they collectively pay just enough to induce the potential entrant to join. While theory does not predict which equilibrium will be played, prospects for international cooperation in the face of assurance games are relatively good (Barrett 2013). The model furthermore contains an option in which only enthusiasts contribute to making side-payments. In that case, the complexity of intra-club bargaining is reduced. On the other hand, it assumes enthusiasts are willing to carry the whole cost of making side-payments, which implies a stronger form of enthusiasm.

Following the Chander–Tulkens rule, the baseline assumption is that costs of side-payments are distributed in proportion to the benefit from expansion $\Delta_{im,}$. As noted, the model contains an option to disallow transfers where the donor's GDP per capita is less than that of the recipient. Then only the benefits to members richer than the potential entrant matter, and only those members share the cost if a deal is struck.

Regarding the division of surplus between the club and the new entrant, I follow Barrett (2001) in assuming the new entrant is paid just enough to cover its net cost of joining. Barrett asks rhetorically whether it would not be fairer that the new entrant receive a share of the surplus created by its accession. He notes, however, that while the new entrant receives no gain from its own accession, it does receive a share of the total gains the club produces by limiting global emissions, and argues that its assessment of the offer's fairness would likely be guided by the total gain.

Negotiations take place sequentially, starting with the candidate providing the largest net benefit.

4.2 Model Steps

The model includes three steps.

Initialization: The modeler chooses the inputs listed in Table 2.

Step 1: Enthusiastic actors join the club *automatically* (as founders).

Step 2: The model calculates the benefit to each member from the entry of each non-member and the compensation each non-member requires to join (subject to the chosen fairness constraints). The model identifies the non-member providing the largest ratio between external benefit provided and private net cost of joining. If this ratio is larger than unity, the non-member joins the club in return for a side-payment. The step is repeated until no more opportunities for mutually beneficial side-payments exist.

Step 3: Enthusiastic actors assess whether they are better off than they would be in the absence of the club. If this is the case, they remain members. If not, they leave the club. If at least one enthusiast leaves and at least one enthusiast stays, new negotiations start between the remaining enthusiast(s) and the reluctant parties by returning to Step 2.

Input	Explanation	Baseline model values	Sensitivity test values
Vulnerability Weight	Degree of differentiation between more and less vulnerable countries	1	0; 2
Global Damage	The difference in global climate damage costs between the business-as-usual (no-club) scenario and the scenario where *all* actors spend 1% of their GDP to mitigate climate change	2% of GGP/ 3% of GGP	
Enthusiasts	Which actors are "enthusiastic"?	See Table 3	

Who pays?	Which members contribute to side-payments?	Enthusiasts/All
Equity	Rules out side-payments where the donor is poorer than the recipient (in terms of GDP per capita).	On/Off

Table 2: Model inputs set at initialization. Note: The first two inputs are empirical parameters with uncertain values, included for the purpose of sensitivity analyses. The last three are characteristics of the negotiation process – and are the analysis's main foci.

5 Results and discussion

5.1 Low returns from cooperation

I begin with the scenario where the global benefit-cost ratio is 2, that is, quite low, and where reluctant members do not contribute to side-payments for further expansion. These assumptions are not very conducive to cooperation, and only two initial coalitions are able to attract enough members to make the club pay off for the founders, compared to BAU. One such coalition consists of the three largest emitters; the other consists of the BRICS countries. Further results are presented in Table 3. The first column lists the different constellations of enthusiasts that are examined. The second column contains the most important outcome variable, namely the percentage of global emissions covered by the club when membership stabilizes. The next columns presents the size of total side-payments, and the last two columns list the country receiving the largest payment and the highest-emitting non-member, respectively. The text below presents also some additional information about the outcomes.

According to the model, the three largest emitters – who themselves account for 50% of global emissions – manage to attract 85 reluctant actors accounting for another 35% of emissions, through side-payments totaling US$ 115 billion, or 0.26% of the three enthusiasts' combined GDP. The largest absolute payment is received by Russia and amounts to US$ 19 billion. The largest payments relative to GDP are, on the other hand, received by the smallest emitters, whose membership costs are compensated nearly 100%.

Japan, Brazil, and Canada together account for half of the emissions not covered by the club. Japan, for example, would require a side-payment of US$ 48 billion to join. As shown in Table A.1 in the Technical Appendix, the side-payment required by a country is a positive function of its GDP and a negative function of its vulnerability and emissions, whereas the side-payment offered to a country is a positive function of its emissions. Large economies with low vulnerability and low emissions intensities are therefore most likely to remain outside the club.

The club formed by the BRICS countries – who account for 42% of emissions – grows to cover 56% of emissions, through side-payments totaling US$ 27 billion, or 0.26% of the three actors' combined GDP. The largest emitters recruited are Indonesia, Iran, and Malaysia. No OECD member joins this club because these countries require larger side-payments than the BRICS countries are willing to pay.

Enthusiasts	Emissions covered	Enthusiasts' expenditure on side-payments (% of GDP)	Enthusiasts' expenditure on side-payments (US$ billions)	Largest side-payment: recipient, (US$ billions)	Highest-emitting non-member
China	0	-	-	-	-
US	0	-	-	-	-
EU	0	-	-	-	-
China, US	0	-	-	-	-
China, EU	0	-	-	-	-
US, EU	0	-	-	-	-
China, US, EU	85.3	0.26	115	Russia, 19	Japan
BASIC	0	-	-	-	-
BRICS	56	0.17	27	Indonesia, 8	US

Table 3: Equilibrium clubs when *Global Damage* is 2 and only enthusiasts offer side-payments

The next scenario assumes reluctant members also contribute to side-payments, which implies that the availability of funds increases as the club grows. This positive feedback may give rise to cascade effects. Table 4 shows that the biggest change in the results is that an initial club consisting of the United States and the European Union will grow to near universal participation and persist. The side-payments involved are quite substantial, with total transactions worth US$ 245 billion. However, because some actors receive side-payments to join but later pay others to do the same, the sum of actors' net side-payments is lower (US$ 158 billion). The two enthusiasts carry the bulk of the load, paying out 0.42% of their combined GDP. The only other actor paying more than it receives is China. It is therefore not surprising that adding China to the list of enthusiasts does not change the outcome much; it only distributes costs more evenly between the top three emitters.

Enthusiasts	Emissions covered	Enthusiasts' expenditure on side-payments (% of GDP)	Total expenditure on side-payments (US$ billions): net; gross	Largest side-payment: recipient; amount (US$ billions)	Highest-emitting non-member
China	0	-	-	-	-
US	0	-	-	-	-
EU	0	-	-	-	-
China, US	0	-	-	-	-
China, EU	0	-	-	-	-
US, EU	95.3	0.42	158; 245	Brazil; 19	Japan
China, US, EU	95.3	0.35	158; 211	Brazil; 19	Japan
BASIC	62.3	0.45	63; 78	Russia; 16	US
BRICS	62.3	0.31	50; 59	Saudi Arabia; 7	US

Table 4: Equilibrium clubs when *Global Damage* is 2 and all members offer side-payments

The BRICS-led coalition is less affected by allowing reluctant members to make side-payments. However, the change means that Russia is no longer needed as an enthusiast, but can be recruited to an initial coalition of the BASICs via side-payments.

When assuming all members can contribute to side-payments, transfers from a poorer to a richer country may occur. While mutually beneficial, such transfers would unlikely be politically acceptable to the poorer country. I therefore tested what happens if transfers can flow only from actors with higher GDP per capita to actors with lower GDP per capita (no table). In this scenario, the EU–US coalition collapses; however adding China to the list of enthusiasts now makes a big difference and results in 85% participation. The BRICS-based coalition persists, but achieves lower participation (53%) than in both previous scenarios. The BASIC-based coalition collapses.

5.2 Larger potential returns to cooperation

Assuming larger returns from abatement expenditure means there are larger potential gains from cooperation. Increasing *Global Damage* from 2 to 3 has a large positive effect on participation. Even when enthusiasts alone make side-payments, all initial coalitions considered persist and grow to cover between half and nearly all of global emissions, as shown in Table 5. The total value of side-payments varies between US$ 40 and 200 billion. It is generally a positive and increasing function of equilibrium club size, but exceptions to this pattern exist, for example when changing Russia's status from reluctant to enthusiastic (i.e., when changing the list of enthusiasts from the BASIC countries to the BRICS countries).

Enthusiasts	Emissions covered (% of global)	Enthusiasts' expenditure on side-payments (% of GDP)	Enthusiasts' expenditure on side-payments (US$ billions)	Largest side-payment: recipient; amount (US$ billions)	Highest-emitting non-member
China	48.2	0.43	40	India; 13	US
US	67.2	0.38	65	Russia; 18	EU
EU	62.7	0.37	66	Russia; 18	US
China, US	75.7	0.40	107	Russia; 18	EU
China, EU	71.7	0.41	111	Russia; 18	US
US, EU	87.0	0.36	128	Russia; 18	Japan
China, US, EU	95.0	0.45	201	Brazil; 21	Japan
BASIC	61.7	0.52	73	Russia; 18	US
BRICS	62.3	0.35	58	Indonesia; 7	US

Table 5: Equilibrium clubs when *Global Damage* is 3 and only enthusiasts offer side-payments

When side-payment costs can be shared among all members, all coalitions considered here reach virtually universal participation, as shown in Table 6. The total value of side-payments required is rather large, particularly in gross terms, in which they amount to around .5% of GGP (US$ 380 billion).

Restricting side-payments to flow down the rich-poor gradient leads to somewhat less optimistic results, with pronounced effects for initial coalitions of developing countries, as could be expected. Participation in a club led by China is halved, while a club led by the United States and the European Union is hardly affected (Table 7).

Enthusiasts	Emissions covered (% of global)	Enthusiasts' expenditure on side-payments (% of GDP)	Total expenditure on side-payments (US$ billions): net; gross	Largest side-payment: recipient; amount (US$ billions)	Highest-emitting non-member
China	99.7	1.87	278; 528	EU; 134	Switzerland
US	99.7	0.73	242; 413	EU; 134	Switzerland
EU	99.7	0.73	246; 391	US; 66	Switzerland
China, US	99.7	0.84	242; 409	EU; 134	Switzerland
China, EU	99.7	0.84	246; 387	US; 66	Switzerland
US, EU	99.7	0.39	192; 272	Japan; 37	Switzerland
China, US, EU	99.7	0.43	192; 268	Japan; 37	Switzerland
BASIC	99.7	1.64	276, 490	EU, 134	Switzerland
BRICS	99.7	1.49	276, 472	EU, 134	Switzerland

Table 6: Equilibrium clubs when *Global Damage* is 3 and all members offer side-payments

Enthusiasts	Emissions covered (% of global)	Enthusiasts' expenditure on side-payments (% of GDP)	Total expenditure on side-payments (US$ billions): net; gross	Largest side-payment: recipient; (US$ billions)	Highest-emitting non-member
China	47.4	0.38	36; 41	India; 12	US
US	73.9	0.39	91; 101	Russia; 16	EU
EU	69.3	0.38	90; 101	Russia; 16	US
China, US	73.9	0.34	91; 98	Russia; 16	EU
China, EU	69.3	0.33	90; 98	Russia; 16	US
US, EU	92.3	0.40	156; 180	Brazil; 20	Japan
China, US, EU	92.3	0.35	156; 176	Brazil; 20	Japan
BASIC	51.3	0.21	30; 32	Indonesia; 6	US
BRICS	56.0	0.19	31; 32	Indonesia; 6	US

Table 7: Equilibrium clubs when *Global Damage* is 3 and all members offer side-payments but only to actors with lower GDP per capita than they themselves have

Decreasing *Global Damage* much below 2 quickly eradicates all clubs. The minimum value for which any club succeeds is 1.9 when only enthusiasts make side-payments and 1.8 when reluctant members also contribute. These are clubs starting with the three largest emitters.

Other notable thresholds are the minimum values of *Global Damage* for which only one or two enthusiasts are needed to initiate a successful club. When only enthusiasts pay, a US–EU coalition persists for values above 2.1. The club grows to cover 82.6% of emissions through side-payments totaling US$ 126 billion. The United States alone can start a club for values above 2.8, which grows to cover 62%

of emissions through side-payments totaling US$ 54 billion.[4] When all member may contribute to side-payments, the minimum values are 1.9 for the United States and European Union together, and 2.4 for the United States alone. These results are summarized in Table 8. These values for the global cost-benefit ratio of mitigation do not appear entirely unrealistic.

Who contributes to side-payments?	Enthusiasts	Minimum benefit-cost ratio for persistence	Emissions covered (% of global)	Net total expenditure on side-payments (US$ billions)
Enthusiasts	China, US, EU	1.9	83.9	105
	US, EU	2.1	82.6	126
	US	2.8	62.0	54
	EU	2.9	62.6	69
All members	China, US, EU	1.8	95.1	157
	US, EU	1.9	95.3	158
	US	2.4	86.3	139
	EU	2.4	99.7	247

Table 8: Minimum benefit-cost ratios for which coalitions persist, and resulting membership and associated side-payments

5.3 Sensitivity to assumed damage-cost distribution

As explained in the model description, relative vulnerabilities of model countries have an empirical basis, but the absolute differences are implemented rather arbitrarily. I therefore check what happens under alternative assumptions about the variance of vulnerabilities, while keeping constant global BAU damage costs. For brevity, this is applied to only one of the scenarios presented above, namely the scenario where *Global MDC* equals 3 and only enthusiasts make side-payments.[5] Table 9 shows the results under the extreme assumption that damage costs are proportional to GDP, that is, when no weight is given to the Notre Dame Vulnerability Index. Relative to the baseline assumption, this implies stronger incentives for climate action by the most developed countries, and lower incentives for less developed countries, including China. China as a lone enthusiast therefore fails to establish a successful club. Apart from that, changes are more marginal. Initial coalitions including the United States or the European Union reach

[4] For the European Union the threshold value is 2.9 (2.87).

[5] This scenario was chosen because it produced the fewest corner solutions among scenarios presented so far.

higher levels of participation and pay larger amounts than under the baseline assumption. The opposite happens for coalitions based on the BRICS and BASIC countries.

Enthusiasts	Emissions covered (% of global)	Enthusiasts' expenditure on side-payments (% of GDP)	Enthusiasts' expenditure on side-payments (US$ billions)	Largest side-payment: recipient; amount (US$ billions)	Highest-emitting non-member
China	0	-	-	-	-
US	72.7	0.63	108	Russia; 19	EU
EU	68.6	0.62	111	Russia; 19	US
China, US	77.9	0.48	127	Russia; 19	EU
China, EU	87.0	0.84	230	US; 102	Japan
US, EU	94.7	0.72	219	Brazil; 22	Japan
China, US, EU	95.3	0.47	209	Brazil; 22	Japan
BASIC	55.7	0.33	46	Russia; 19	US
BRICS	59.2	0.26	43	Indonesia, 8	US

Table 9: Equilibrium clubs when *Global Damage* is 3, only enthusiasts offer side-payments, and *Vulnerability Weight* is 0

Table 10 shows the results when *Vulnerability Weight* is set to 2, that is, assuming vulnerable actors bear a larger share of damage costs than in the baseline scenarios. The implied damage-cost distribution is shown in Figure 2. This is also a rather extreme assumption, implying that China has an incentive for unilateral mitigation, to avoid the damages to itself caused by own emissions.[6] The European Union and the United States, in contrast, have little incentive for action. Notably, these actors would leave an initial coalition between one of them and China, who on its own succeeds in attracting enough other members for the club to persist.

Enthusiasts	Emissions covered (% of global)	Enthusiasts' expenditure on side-payments (% of GDP)	Enthusiasts' expenditure on side-payments (US$ billions)	Largest side-payment: recipient; amount (US$ billions)	Highest-emitting non-member
China	53.2	0.59	56	Russia; 18	US

[6] If the cost of making side-payments can be shared with reluctant members, enthusiasts are actually not necessary for club success under this damage-cost distribution assumption. A club would start with China's unilateral action and grow to virtual universal coverage through side-payments totaling US$ 324 billion. China would be spending nearly 2% of its GDP on side-payments.

US	0[7]	-	-	-	-
EU	0[a]	-	-	-	-
China, US	53.2	0.59	56	Russia; 18	US
China, EU	53.2	0.59	56	Russia; 18	US
US, EU	82.6	0.25	89	Russia; 18	Japan
China, US, EU	89.4	0.32	141	Russia; 18	Japan
BASIC	62.5	0.54	76	Russia; 18	US
BRICS	64.1	0.43	69	Saudi Arabia; 7	US

Table 10: Equilibrium clubs when *Global Damage* is 3, only enthusiasts offer side-payments, and *Vulnerability Weigh*t is 2

The sensitivity analysis shows that side-payments are effective under a wide range of assumptions about the distribution of damage costs. However, it also shows that which specific clubs emerge depends on how damage costs are distributed.

5.4 The importance of asymmetry

Asymmetry between actors is an important enabling condition for side-payments to work. If all model actors were assumed equal in GDP, emissions, and vulnerability, then for any of the considered coalitions to succeed, the threshold value of Global MDC would have to be 34 (!),compared with 1.8 when empirical distributions are used. This result supports the analysis of Barrett (2001). It also indicates that models assuming homogenous actors sometimes provide predictions that are misleading when extrapolated to real-world climate negotiation, where strong asymmetries exist.

6 Concluding remarks

This paper has presented a formal model of climate clubs and investigated the potential effect of side-payments. The model assumes that one or a small group of actors are willing to lead through reducing emissions and offering side-payments to others who follow suit. The model incorporates empirical data on countries' GDP, emissions, population, and vulnerabilities to climate change. It allows for different

[7] China mitigates unilaterally.

assumptions about which countries are willing to lead, how to share the cost of side-payments as the club grows, equity considerations, the benefit-cost ratio of mitigation, and the distribution of climate damage costs.

The results may be summarized as follows:

1. A single large economy can use side-payments to attract actors responsible for a substantial share of emissions, given that the global benefit-cost ratio of abatement is at least around 3. Some tens of billions of US dollars would be needed annually. All actors would be better off relative to BAU.
2. If two or three large emitters enthusiastically join forces, they would have the ability to incentivize a large number of countries to follow suit.
3. If new members contribute to funding side-payments, virtually universal participation can be ensured for moderate abatement cost-benefit ratios. However, in large clubs, the coordination problem among members would certainly be non-trivial. Side-payments in the hundreds of millions of US dollars would be needed.
4. Large economies like Russia, Brazil, India, Saudi Arabia, and Indonesia would receive the largest side-payments in absolute terms. However, small emitters receive the largest compensation relative to their GDP.
5. If the richest countries are reluctant, it will be difficult to get them on board through side-payments even if transfers to rich countries are not ruled out on grounds of fairness.

Overall, the results present a rather optimistic picture of the potential for clubs based on a few or even a single enthusiast to grow through side-payments. In comparison, Author et al. (forthcoming) found that either exclusive club goods or conditional mitigation pledges are effective in isolation only under somewhat more limited conditions. (However, in combination they are effective under a quite broad set of conditions.) The model therefore indicates that of the instruments assessed, side-payments provide the most leverage for enthusiasts to recruit reluctant actors into a climate club. One reason for their relative effectiveness is that they are targeted at potential entrants only. In contrast, club benefits accrue also to existing members, and conditional mitigation efforts benefit both members and non-members alike.

Political feasibility is a major obstacle for side-payments. The required side-payment figures would likely be politically extremely challenging to muster. However, developed countries have already committed to a goal of jointly mobilizing US$ 100 billion annually by 2020 (UNFCCC 2009). Whether this target is achievable can be debated; nevertheless, this paper suggests that such a figure could buy substantial climate cooperation under a broad range of assumptions if payments were made conditional on mitigation by recipients. Exogenous conflicts between countries may also obstruct some of the deals that are successful in the model. For example, the model indicates that China and the United States working together could give rise to a quite effective climate club through using mutually beneficial side-payments. When the list of countries receiving payments includes India, Russia, Iran, and Saudi Arabia, enthusiasm will be put to a hard test. The political feasibility of side-payments may be increased by using other forms than cash transfers, for example technology transfer or emissions trading (Victor 2011; McGinty 2014). However, if the objective of compensation is blurred by other overlapping objectives, theoretical efficiency is reduced (Finus 2003). More research is needed to answer which side-payment form best balances political feasibility and theoretical efficiency.

This paper’s results are based on the assumption that actors comply with their commitments. As discussed in the literature review, enforcing donor compliance will likely be a significant challenge.

Multilateral funds may help in this respect by serving as a commitment device, whereby donors commit by placing money into a fund that pays out money to recipients upon delivery of abatement efforts. There seems, therefore, to be a theoretic role for structures like the Green Climate Fund, which was established after Copenhagen. Whether that fund is likely to serve such a role in practice will not be explored here.

8 Bibliography

Altamirano-Cabrera, J.-C., & Finus, M. (2006). Permit trading and stability of international climate agreements. *Journal of Applied Economics, 9*(1), 19–47.

Andresen, S. (2014). Exclusive Approaches to Climate Governance: More Effective than the UNFCCC? In T. L. Cherry, J. Hovi, & D. M. McEvoy (Eds.), *Toward a New Climate Agreement* (pp. 155–166). London: Routledge.

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

Barrett, S. (1997). The strategy of trade sanctions in international environmental agreements. *Resource and Energy Economics, 19*(4), 345–361.

Barrett, S. (1999). The credibility of trade sanctions in international environmental agreements. *World bank discussion papers*, 161-172.

Barrett, S. (2001). International cooperation for sale. *European Economic Review, 45*(10), 1835–1850.

Barrett, S. (2013). Climate treaties and approaching catastrophes. *Journal of Environmental Economics and Management, 66*(2), 235–250.

Botteon, M., & Carraro, C. (1997). Environmental coalitions with heterogeneous countries: Burden-sharing and carbon leakage. *Fondazione Eni Enrico Mattei Working Paper*(24.98).

Chander, P., & Tulkens, H. (1995). A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. *International tax and public finance, 2*(2), 279–293.

d'Aspremont, C., Jacquemin, A., Gabszewicz, J. J., & Weymark, J. A. (1983). On the stability of collusive price leadership. *Canadian Journal of economics, XVI*(1), 17–25.

Falkner, R. (2015). A minilateral solution for global climate change? On bargaining efficiency, club benefits and international legitimacy. Grantham Research Institute on Climate Change and the Environment. Working Paper No. 197.

Falkner, R., Stephan, H., & Vogler, J. (2010). International climate policy after Copenhagen: Towards a 'building blocks' approach. *Global Policy, 1*(3), 252–262.

Finus, M. (2003). Stability and design of international environmental agreements: the case of transboundary pollution. In T. Tietenberg, H., & H. Folmer (Eds.), *International yearbook of environmental and resource economics* (pp. 82–158, Vol. 2003/2004). Cheltenham, UK: Edward Elgar.

Finus, M., Van Ierland, E., & Dellink, R. (2006). Stability of climate coalitions in a cartel formation game. *Economics of Governance, 7*(3), 271–291.

Fuentes-Albero, C., & Rubio, S. J. (2010). Can international environmental cooperation be bought? *European Journal of Operational Research, 202*(1), 255–264.

Fujita, T. (2006). A comment on "International Cooperation for Sale". *Economics Bulletin, 8*(1).

Glanemann, N. (2012). Can international environmental cooperation be bought: Comment. *European Journal of Operational Research, 216*(3), 697–699.

Global Carbon Project (2014). Fossil fuel and cement emissions 2013. Retrieved 1 October 2014 from: http://www.globalcarbonatlas.org/?q=en/emissions.

Goodell, J. (2014). The Secret Deal to Save the Planet: Inside the high-stakes drama behind Obama's China climate talks. *Rolling Stone, December 9*.

Helm, D., Hepburn, C., & Ruta, G. (2012). Trade, climate change and the political game theory of border carbon adjustments. *Oxford Review of Economic Policy, Autumn, 28*(2), 368–394.

Kroll, S., & Shogren, J. F. (2008). Domestic politics and climate change: international public goods in two-level games. *Cambridge Review of International Affairs, 21*(4), 563-583.

Mayer, F. W. (1992). Managing domestic differences in international negotiations: the strategic use of internal side-payments. *International Organization, 46*(4), 793-818.

McGinty, M. (2007). International environmental agreements among asymmetric nations. *Oxford Economic Papers, 59*(1), 45–62.

McGinty, M. (2010). International environmental agreements as evolutionary games. *Environmental and Resource Economics, 45*(2), 251–269.

McGinty, M. (2011). A Risk-Dominant Allocation: Maximizing Coalition Stability. *Journal of Public Economic Theory, 13*(2), 311–325.

McGinty, M. (2014). Improving the design of international environmental agreements. In T. L. Cherry, J. Hovi, & D. M. McEvoy (Eds.), *Toward a new climate agreement. Conflict resolution and governance* (pp. 128–142). London and New York: Routledge.

Nagashima, M., Dellink, R., van Ierland, E., & Weikard, H. P. (2009). Stability of international climate coalitions – A comparison of transfer schemes. *Ecological Economics, 68*(5), 1476–1487.

Neumayer, E. (2001). *Greening trade and investment: environmental protection without protectionism*: Earthscan.

Nordhaus, W. D. (2015). Climate Clubs: Designing a Mechanism to Overcome Free-riding in International Climate Policy. *American Economic Review, 105*(4), 1–32.

Notre Dame Global Adaptation Index (2014). Vulnerability scores for 2012. Retrieved 7 April 2014 from http://index.gain.org/about/download.

Richter, A., & Grasman, J. (2013). The transmission of sustainable harvesting norms when agents are conditionally cooperative. *Ecological Economics, 93*, 202–209, doi:http://dx.doi.org/10.1016/j.ecolecon.2013.05.013.

Sprinz, D., & Vaahtoranta, T. (1994). The interest-based explanation of international environmental policy. *International Organization, 48*(01), 77-105.

Stewart, R. B., Oppenheimer, M., & Rudyk, B. (2013). Building blocks for global climate protection. *Stanford Environmental Law Journal, 32*(2), 12–43.

Underdal, A., Hovi, J., Kallbekken, S., & Skodvin, T. (2012). Can conditional commitments break the climate change negotiations deadlock? *International Political Science Review, 33*(4), 475-493.

UNFCCC (2009). Decicions 2/CP.15: The Copenhagen Accord. Available from: http://unfccc.int/resource/docs/2009/cop15/eng/11a01.pdf#page=4

Urpelainen, J. (2013a). Can strategic technology development improve climate cooperation? A game-theoretic analysis. *Mitigation and Adaptation Strategies for Global Change, 18*(6), 785–800.

Urpelainen, J. (2013b). A model of dynamic climate governance: dream big, win small. *International Environmental Agreements: Politics, Law and Economics, 13*(2), 107–125.

Victor, D. G. (2011). *Global warming gridlock: creating more effective strategies for protecting the planet*. Cambridge: Cambridge University Press.

Weidmann, N. B., & Salehyan, I. (2013). Violence and ethnic segregation: a computational model applied to Baghdad. *International Studies Quarterly, 57*(1), 52–64.

Weikard, H.-P., Finus, M., & Altamirano-Cabrera, J.-C. (2006). The impact of surplus sharing on the stability of international climate agreements. *Oxford Economic Papers, 58*(2), 209–232.

Weischer, L., Morgan, J., & Patel, M. (2012). Climate Clubs: Can Small Groups of Countries Make a Big Difference in Addressing Climate Change? *Review of European Community and International Law, 21*(3), 177–192.

World Bank (2014a). GDP 2013 (constant 2005 US$). Retrieved 1 October 2014, from http://databank.worldbank.org/data/views/variableselection/selectvariables.aspx?source=world-development-indicators#s_cbudsjet.

World Bank (2014b). Population, total. Retrieved 1 October 2014, from http://databank.worldbank.org/data/views/variableselection/selectvariables.aspx?source=world-development-indicators#s_cbudsjet.

World Resources Institute (2014). CAIT 2.0 Total GHG emissions from land-use change and forestry 2011. Retrieved 1 October 2014 from: http://cait2.wri.org/wri/Country%20GHG%20Emissions.

Technical Appendix

The following provides a technical and complete description of the model expressed through equations and pseudo-code. For further explanation, please see the verbal description in the main manuscript. The model code is available from the authors.

Variable or parameter	Explanation	Measurement units or options
	Global input parameters	
Global Damage	The difference in global climate damage costs between the no-club scenario and the scenario where *all* actors spend 1% of their GDP to mitigate climate change. See Table 2 for numerical values.	% of Gross Global Product (GGP)
Vulnerability Weight	Degree of differentiation between more and less vulnerable countries. See Table 2 for numerical values.	-
Club fee	The abatement cost for members. Fixed at 1.	% of GDP
Donors	The set of agents who potentially make side-payments. The set is dynamic	Enthusiasts; all members
Equity	Option to rule out side-payments where the donor is poorer than the recipient (in terms of GDP per capita).	On; Off
	Agent-specific input parameters	
GDP_i	GDP at market exchange rates, 2013 (World Bank 2014a).	Share of GGP
Pop_i	Population, 2013 (World Bank 2014b).	-
$Emissions_i$	Emissions from fossil fuels and cement, 2013 (Global Carbon Project 2014) + Net emissions from land use change and forestry, 2011 (World Resources Institute 2014).	Share of global emissions
$NDGAIN_i$	Vulnerability Scores, 2012 (Notre Dame Global Adaptation Index 2014). ∈[0,1]	-
$Enthusiast_i$	Is *i* "enthusiastic"?	Yes; No
	Agent-specific variables derived in the model	
$Vulnerability_i$	The percentage loss in GDP_i arising when the global loss is 1% of GGP. $\frac{NDGAIN_i + (Vulnerability\ Weight - 1) \times (NDGAIN_i - \overline{NDGAIN})}{\overline{NDGAIN}}$ $\overline{NDGAIN}$ is GDP-weighted mean *NDGAIN.*	$\frac{\%\text{ of GDP}_i}{\%\text{ of GGP}}$
$Damage\ cost_i$	The difference in *i's* damage costs between the no-club scenario and the scenario where *all* actors spend 1% of their GDP to mitigate climate change. $Vulnerability_i \times Global\ damage$	% of GDP_i
WTA_i	Side-payment that would compensate non-member *i*'s net cost of joining. $(Club\ fee - Damage\ cost_i \times Emissions_i) \times GDP_i$	% of GGP
WTP_{im}	Benefit to *m* when *i* joins. $Damage\ cost_m \times Emissions_i$	% of GDP_m
$Ratio_i$	The ratio between the sum of benefits to donors if *i* joins and the side-payment needed to compensate *i*'s net cost of joining.	-

	$\sum_{m \neq i}^{Donors} (WTP_{im} \times GDP_m)/WTA_i$	
Side-payment$_{kl}$	Payment incurred by *l* to induce *k to* join[8]. $WTP_{kl}/Ratio_k$	% of GDP_l
Payoff$_i$	*Payoff* is normalized to be (- *Damage cost*$_i$) in the BAU, and (-*Club fee*) if all countries are club members. *Payoff* for members equals the difference in *i's* damage costs between the realized scenario and universal club scenario minus abatement costs plus the sum of side-payments received minus the sum of side-payments paid. For non-members, only the first term enters. $-Damage\ cost_i \times \left(\sum_{n}^{non-members} Emissions_n\right) - Club\ fee + \sum_{m \neq i}^{members} Side-payment_{im} - \sum_{m \neq i}^{members} Side-payment_{mi}$	% of GDP_i

Table A.1: Model parameters and variables

Initialization

Each agent *i* (in random order):
    Calculate *Vulnerability*$_i$
    Calculate *Damage cost*$_i$

Execution

Step 1

Each enthusiast (in random order):
    Become member

Step 2

Loop until no more agents want to join:
    Each non-member *i* (in random order):
        Calculate WTA_i
        If *Equity* = off:
            Ask each Donor *m* (in random order):
                Calculate WTP_{im}
        If E*quity* = on:
            Ask each Donor *m* whose $GDP_m/Population_m > GDP_i/Population_i$(in random order):
                Calculate WTP_{im}

[8] The additional mitigation by the club is distributed proportional to WTP_{im}. Dividing by the potential entrant's benefit-cost ratio implies that the members undertake the minimum additional effort necessary to compensate the potential entrants cost of membership.

Calculate $Ratio_i$

Non-member with the largest $Ratio_i$

If $Ratio_i > 1$:

Become member

If *Equity* = off:

Ask other members *m* (in random order):

Pay $Side\text{-}payment_{ml}$

If *Equity* = on:

Ask other members *m* whose $GDP_m/Population_m > GDP_i/Population_i$ (in random order):

Pay $Side\text{-}payment_{ml}$

End loop

Step 3

Loop until no more enthusiasts want to leave the club:

Each enthusiast *i* (in random order):

Calculate $Payoff_i$

If $Payoff_i < -$ *Damage Costs*[9] $_i$:

Become non-member

Reluctant members (in random order):

Become non-members

End loop

Go back to Step 2

[9] The right-hand side of the inequality is payoff in the no-club scenario.