Climate Clubs, Participation and Efficiency: Can "Reparations" Help?

Prajit K. Dutta* and Haaris Mateen**
Columbia University and University of Houston*

November 1, 2024

Abstract

The paper studies how international transfers can Pareto improve agreements on emission reductions. It re-interprets as a club model a well-studied model of treaty participation. Without transfers, equilibrium club size is limited to being at most 2 or 3 countries. With bi-lateral transfers, transfers voluntarily given by club members to non-members, club size increases to at least $\frac{N}{2}$, where N is the total number of countries. Furthermore, inclusive of transfers, there is a Pareto improvement for all N countries. Any further expansion and welfare improvement requires multi-lateral transfers in which non-members can also commit to making transfers. In that setting, there is an equilibrium with full participation of all N countries and Utilitarian Pareto optimal emissions reductions by every country. If the transfer choices are sequentially made, then that globally implemented reduction is the unique equilibrium action.

1 Introduction

Climate Change is a global (dynamic) externality. The sovereignty of countries and the absence of relevant international courts means that agreements or treaties to reduce emissions must be incentive compatible (IC) for each country. Past environmental agreements have shown the consequences of ignoring these incentive compatibility constraints. For example, the Kyoto Global Climate Agreement in 1997 was touted to have "globally agreed emissions targets." The agreement was not ratified by the U.S. Similarly, the recent Paris Climate Accords set up self-agreed targets. However, these targets have not been achieved, and there has been a lot of free-riding.

The concept of Climate Clubs has been offered as one possible solution to this problem by Nordhaus (2015). A starting point is an acknowledgment that

---

crafting a treaty that is incentive compatible for every nation is a tall order as is monitoring such an agreement. Instead, Nordhaus proposed that a mini multilateral agreement might be an achievable target.

The proposal is an adaptation of the idea of clubs first introduced by Buchanan (1965). He had suggested that it might be easier to have a small group of individuals come together to provide a public good, i.e., for a small group to form a club. A small group - Buchanan argued - would be able to monitor each other better and set up requisite entry barriers. The Nordhaus (2015) extension of this - in the climate change context - is that a club of possibly a small number of nations may be able to agree to set and enforce club emission targets such that no IC constraints would need to be checked for club members.

One key requirement of Buchanan's club is for the good in question to be (non-rival but) excludable. Hence, non-members can be excluded from the benefits generated by the actions of club members (unless they pay the requisite fee and join the club). This feature is evidently not true in the climate context. If club members cut their emissions thereby lowering the stock of greenhouse gases (GHGs), then non-members also benefit. To address this issue, Nordhaus (2015) proposed that non-members be incentivized to "fall in line" and cut emissions under the threat of trade sanctions, i.e., tariffs. As explained in Nordhaus (2015) page 1341, "nonparticipants are penalized...[through] uniform percentage tariffs on imports of nonparticipants into the club region."

While the idea of imposing tariffs is attractive to many people - including a certain US Presidential candidate - critics have pointed out that it violates the GATT agreement that potential club members are signatories to. For example, the Most Favored Nation (MFN) article of GATT (WTO 1994) states that, "Under the WTO agreements, countries cannot normally discriminate between their trading partners. Grant someone a special favor...and you have to do the same for other WTO members." More recently, countries in the EU have tried to implement cross border adjustment mechanisms that try to go around this problem by imposing tariffs on any import that doesn't meet carbon abatement targets. The rules have invited threats of retaliation from trading partners, as well as protests from domestic manufacturers that are threatened by the policy.[1]

In this paper, we ask a different question – rather than the threat of sanctions, is it possible to give positive incentives to push countries to reduce emissions? This idea is rooted in the Coasian Perspective, that any externality can be ameliorated by side-payments. In particular, a developing country may reduce emissions if paid to do so. Our paper is the first game theoretic formulation of the club model, albeit with a focus not on tariffs but rather on transfers.

There are at least two reasons to think about this question. First, it can be motivated by a financing concern; that developing economies will be simply unable to marshall the resources, either by domestic fiscal means or through bor-

---

[1] From a modeling and analysis perspective, it is also a challenge to find a tractable yet detailed enough trade model that can be merged with a climate change framework. Even without the latter complication, as we have seen from a discussion of the Trump tariffs, there is no consensus on how to determine the effects of tariffs. Different models suggest anything from significant positive impacts on the nation imposing them to significant negative ones.

rowing on international markets, to make the required large-scale low emission transition. Or, second, the reparations argument; that developing economies are being asked to bear the brunt of a problem that they did not create and that is unfair.

Either way, as economists we can focus on the Coasian argument - that if developing economies reduce their emissions due to side-payments that can potentially lead to a Pareto improvement for all countries. In this way, payments - that we will interchangeably also call transfers - are different from tariffs. Tariffs are unlikely to lead to Pareto improvements unless they are never used; unless, merely their threat suffices to change emissions behavior. By contrast, transfers even when used, can improve welfare for both donor and recipients. If the donors pay x to recipients to alter their choice from A to B and that change improves donors' welfare by $y > x$, then both parties are better off.

Indeed, international payments have emerged as the single most contentious issue at recent COP discussions. Historically, developing countries have repeatedly pushed for payments - justifying them as reparations - and developed countries have pushed back for fear of opening a Pandora's Box of historical claims.

A breakthrough happened at COP 27 where even the United States finally signed on to the idea of a climate fund. The New York Times thusly reported on Nov 19, 2022: *"In a First, Rich Countries Agree to Pay for Climate Damages in Poor Nations"*. In the Report of the Conference it was stated that "Governments took the ground-breaking decision to establish new funding arrangements, as well as a dedicated (Loss and Damage) Fund, to assist developing countries in responding to loss and damage." It was further acknowledged that "transformation to a low-carbon economy will require investments of at least USD 4-6 trillion a year."

However, many questions remained: who would pay, how much, when, should larger players like China, that are now able to pay but were historically not high emitters, receive payments? Etc.[2]

Whilst the discussion around international transfers has been vigorous at various international forums, it has almost always been seen as a zero-sum fight - developed countries perceive that they will be the losers by funding developing ones. The latter see it as a rectification of past wrongs, reparations in the same spirit as "paying for" the crimes of slavery or colonialism etc. As argued above, with the back context of Coase, economists see it as a non-zero sum issue, that there is a potential for both developed and developing countries to be better off if the latter cut emissions and transition to renewables. The question then becomes a cost-benefit issue for developed countries: is the benefit from developing countries cutting emissions greater than the cost of funding those cuts? To answer that question we need an understanding of how the Coasian

---

[2] There has been a back-slide subsequently. As the Washington Post reported on October 9, 2023 - "Promises from some of the world's biggest economies, including the United States and China, haven't been panning out. Many are years behind schedule or still years away from sending money." Only $9.3 billion dollars had been sent to the Green Climate Fund by then versus the estimated trillions required.

mechanism will work.

That is what we propose to do in this paper. Naturally, our model will have many simplifications and should be seen as a first cut at the question. It is, however, a question that has been looked at almost not at all so far.[3]

The paper is divided into two halves, united in their focus on the effect of transfers on equilibrium treaties. We start with a baseline well-studied simple model of treaty building, Barrett (2005). In that model, countries decide whether or not to sign onto a treaty that requires them to commit to cutting emissions. These cuts help signatories since emissions are a public bad but also help non-signatories who get to free-ride on the cuts. The focus in the model is on participation: how many countries would sign such a treaty? There is a vast literature studying that model and Barrett (2003, 2005) is a good summary. The general conclusion is that treaty participation is extremely limited.

Formally, call the number of willing signatories to a club as K in a world of N countries. Barrett (Chapter 28, 2006) and others show that, in equilibrium, $K \leq 3$, independent of N.[4] This is because there is a discrete jump in payoffs to a country when it opts out of a treaty (and free-rides on the emission cuts of remaining club members). This is analogous to the discrete jump in market share and hence payoffs when a firm makes a very small price cut in a Bertrand competition model. Naturally, that is a disappointing result; sadly, it has been found to be robust to variations in the payoff functions.

We re-interpret the treaty model as a club participation model by adding incentives that could alter the amount of free-riding by non-participants. In the treaty model, there is nothing to incentivize non participants; neither negative incentives like Nordhaus tarrifs nor positive ones like international transfers. As argued above, tariffs might violate GATT and are also difficult to model. Hence, we instead add positive incentives to the model and ask: can one achieve larger participation with (bilateral) transfers that flow from club members to non-members?

The transfers have to be incentive compatible at club level in the sense that a club can decide whether or not to offer them. We analyze in detail a simple transfer scheme that a club might adopt which we call a threshold scheme. In that scheme, transfers are offered if the club size is above a threshold K and - when offered - the size of transfers is optimal for club members. We characterize the size of the club K that can be achieved through such a scheme and show that $K \geq \frac{N}{2}$. We then go on to analyze any transfer scheme that a club might adopt and show the maximal (and minimal) reaches in participation on account of different transfer schemes.

For all transfer schemes studied, there is Pareto improvement relative to the no transfers outcome. Naturally, non-members are at least as well off when they are offered transfers. What is striking is that club members are strictly

---

[3] There are relatively very few papers that have looked at the effect of international transfers in a climate change context. These are discussed in Section 6, and include Carrao-Siniscalco (1993), Chander-Tulkens (1995, 1997), Barrett (2003, 2005) and Dutta-Radner (2023).

[4] The exact variant of the model can be different but the general result is that K is very small, or that N is very small.

better off than in the no transfers case. This occurs because there is a global increase in abatement; club members abate more given that the club is larger and non-members abate more incentivized by transfers.

The baseline treaty model assumes symmetry, i.e., that countries have the same benefit and cost parameters for emission reduction. Naturally, this is a gross simplification. In particular, it does not allow us to ask: which are the countries that are going to sign a treaty (and be members of the club)?

To address that issue, we introduce cost heterogeneity (to reducing emissions). We show that countries that have the highest costs to reducing emissions are most likely to be part of the club, i.e., those are the countries with the greatest incentives to give transfers to have others reduce emissions. To the extent that developed countries have the highest labor and other costs, one can interpret this result as suggesting that transfers will flow from them to developing economies.

Whilst having $\frac{N}{2}$ signatories is a significant improvement on having only three participants, it is still way short of a global treaty that all countries sign onto. Given the scale of the climate change problem, it remains an interesting question as to how can we organize an agreement that will appeal to every country and produce a globally efficient outcome.

In the second half of the paper we show we can indeed do better. The crucial idea here is that we allow *any* country to make transfers (to any other country) rather than restrict transfers to only club members. After all we are dealing with a global externality so a non-club member, country I, is affected by another non-member $J's$ abatement (and it is also affected by what club members K do). So, why not allow I to make payments to J and K? Whereas in the first half of the paper only the K club members make (bilateral) transfers to all other non-members I and J, now we analyze multi-lateral transfers that flow in all directions. Potentially, some transfers would net out. For instance, there might be very little net flowing from one developing country to another but typically it would not be zero, as the (bilateral transfer) club model would require.

Beyond this change, the model remains the same. Each country has to decide whether or not to join a club. Once membership is decided on, in a second stage, the club - acting as one entity - and every non-member acting for itself, decides on a best response transfer to make to others. Finally, in response to those announced transfers, the club - acting as one entity - and every non-member acting for itself, decides on a best response emission reduction.

We show that our setup is a special case of Bernheim-Whinston (1986), i.e., our model is a Common Agency in the sense of that paper. We then directly show that there is an efficient equilibrium, i.e., one in which the globally optimal emission is chosen by every country having been incentivized by appropriate transfers chosen in the second stage. Finally, in that equilibrium, all countries sign on to be in the club, i.e., $K = N$.

However, as is well-known from Bernheim-Whinston (1986), in the Common Agency model there might also be inefficient equilibria since transfers are chosen simultaneously before emissions. This simultaneity can introduce free-riding; both I and J are affected by the emissions chosen by K and are willing to pay

5

to lower that but both prefer to have the other make the payment. When the payment/transfer commitments are simultaneously chosen, I and J can end up choosing inefficiently small transfers consequently.

In the last section of the paper we address that possibility by investigating sequential transfer choices in the second stage: I announces its commitment to K first which is seen by J who then announces its commitment. Otherwise, the model remains unchanged. Each country has to decide whether or not to join a club in the first stage. Once membership is decided on, the club decides on transfers to give non-members and every non-member decides likewise. This stage is sequential. Finally, in response to those announced transfers, the club - acting as one entity - and every non-member acting for itself, decides on a best response emission reduction. We show that in this setting, there is a unique equilibrium emission reduction and that is one in which every country chooses the globally optimal one.

The **Summary of Results** is in the table below. In equilibrium, K is the club size, q_K is the amount of abatement done by each club member, q^* that done by non-members and $\widehat{q}$ is the Utilitarian Pareto Optimum (UPO) abatement that is globally efficient for a world of N countries. Turns out that the abatement done by non-members without transfers is the smallest amount and let us normalize that to 1. For *No Transfer* and *Multilateral (Simultaneous)* modes, the best equilibrium in terms of abatement is reported while for the *Bilateral*, a near-best equilibrium is reported.[5] The last one - *Muliti-Lateral Sequential* - has a unique abatement profile in equilibrum.

Transfer Mode	Club size	Abatement Levels	Total Abatement
None	$K = \{2, 3\}$	$q^* = 1, q_K = K$	$9 + (N - 3)$
Bilateral	$K = \frac{N}{2}$	$q^* = K + 1, q_K = K$	$\frac{N}{2}(N + 1)$
Multilateral (Simultaneous)	$K = N$	$q_K = \widehat{q} = N$	N^2
Multilateral (Sequential)	Variable	$q^* = q_K = \widehat{q} = N$	N^2

Section 2 presents the baseline Treaty Participation (Barrett) model that has no transfers and reports the result on limited participation. Section 3 expands the model to include transfers. It also presents results on expanded treaty participation, Pareto improvement vis-a-vis the baseline model and the differential incentives under heterogenity. Section 4 introduces the multi-lateral model with transfers and presents a full characterization of equilibrium emissions. Section 5 studies sequential transfers in the multi-lateral model. A discussion of the literature and open questions is in Section 6.

2   Clubs Without Transfers

We first present the baseline Barrett model which has no transfers. It is a static model with no heterogenity in payoffs.

---

[5] Meaning that in the Bilateral Transfers case, there is a range of possible club sizes. The largest is somewhat greater than the reported $\frac{N}{2}$.

2.1 Model

There are N countries ($N \geq 2$). Each country i can choose the quantity of abatement $q_i \geqslant 0$. Abatement level q_i can be interpreted as bringing emissions down from a maximum level - say $\bar{e}$ - down to $\bar{e} - q_i$. Abatement contributes to a public good, the sum of all abatements, but it imposes a private cost on country i. In the model, the benefit is linear while the cost is quadratic. Writing q as the vector of abatements, the payoff π_i is

$$\pi_i(q) = b \sum_{k \in N} q_k - \frac{1}{2} c q_i^2 \tag{1}$$

where b and c are non-negative constants that are the same across countries.

Countries can create a club of size $K \leq N$. Club countries abate as a group and therefore internalize any externality between them. This is an assumption - that being in a club allows members to monitor themselves thereby "loosening" incentive constraints (IC) within that sub-group.

Consequently, since the club acts as a synthetic player, it seems natural to take the amount of abatement q^K done by each club member to be the sub-group optimal one, i.e., to assume q^K as $\mathrm{argmax} \sum_{i \in K} \pi_i(q)$.[6] The $N - K$ non-members are, however, free to individualistically choose their abatement levels. Call the best response level, computed from Eq. 1 for each non-member, q^*.[7] It is easy to check that $q^* < q^K$ since the club internalizes the externality that every member's abatement implies for the $K - 1$ others and hence abates more.

Timing - There are two stages: In Stage 1, countries simultaneously decide whether or not to join a Club. Those decisions determine the size of K. Then, in Stage 2, countries simultaneously pick q^K and q^*. Only the latter is a best response while the former is part of the (implicit) contractual requirement when a country joins the club.

Equilibrium - A club size K is an equilibrium if no member wants to drop out and no non-member wants to come in. So, if a country anticipates that $K - 1$ others are going to join the club, it can a) join and abate at q^K (when $N - K$ others are abating at q^*) or b) stay out and abate at q^* (alongside $N - K$ others) and let $K - 1$ countries abate at q^{K-1}. The participation IC for members requires that a) has a higher payoff than b).

The second IC is that non-members do not want to join: abating at q^* (along with $N - K - 1$ others) is better than abating at q^{K+1} alongside the other K club members.

The question of interest is - what is the highest value of K satisfying both IC?

---

[6] Note that due to the linearity of π_i, this argmax is independent of the abatement levels chosen by non-members.

[7] Again, due to the linearity of π_i, q^* is independent of the abatement levels chosen by club members.

2.2 Limited Participation Result

A little bit of computation incorporating the values[8] q^* and q^K leads to the following payoff for a K club member

$$\Pi^K = B(N, K) - CK^2, \tag{2}$$

where $B(N, K) = \frac{b^2}{c}[K^2 + N - K], C = \frac{b^2}{2c}$. Similarly, the payoff for a non-member is

$$\pi^K = B(N, K) - C. \tag{3}$$

Note that a non-member has a higher payoff. This follows straightforwardly from the fact that the public good is identical for members and non-members but the latter abate at a lower level and hence incur a smaller cost.

What would make K^* an equilibrium club size? The first IC to check is that no member wants to leave the club, i.e.,

$$\Pi^{K^*} \geqslant \pi^{K^*-1}.$$

A little bit of algebra on Eqs. 2 and 3 gives us that $K^* \leqslant 3$.

The second IC is that no non-member wants to come into the club, i.e., $\pi^{K^*} \geqslant \Pi^{K^*+1}$. It is not difficult to check that this constraint is not binding since it is satisfied by $K^* \geqslant 2$. That yields the following result in the baseline club model with no incentives targeted at non-members:

Proposition 1 *The largest club size in the model without transfers is that of three countries.*

3 Clubs With Bi-Lateral Transfers

3.1 Model

We retain the base payoffs and abatement possibilities of the baseline Barrett model. There is, again, a first stage in which countries choose whether or not to be in the club and a final stage in which non-members choose their best response abatement while club members abate at the contractually required level q^K, given club size K.

We add to the baseline model an intermediate stage - between the two stages of choosing club membership (or not) and abating, there is a second stage in which club members choose a transfer to make non-members abate more.

Note that, when offered a transfer schedule $\theta(\bullet)$, a non-member's payoff is

$$\pi_i(q; \theta) = b \sum_{k \in N} q_k - \frac{1}{2} c q_i^2 + \theta(q_i), \tag{4}$$

---

[8] Given the linear-quadratic payoffs, it can be shown that $q^K = \frac{Kb}{c}$ and $q^* = \frac{b}{c}$.

while that of a club member is

$$\Pi_i(q;\theta) = b\sum_{k\in N} q_k - \frac{1}{2}cq_i^2 - \frac{N-K}{K}\theta(q_i), \tag{5}$$

where $\theta(q_i)$ is the transfer that each of the $N-K$ non-members receive from a club with K members.

Timing - To summarize, there are three stages. In the first stage, a set of countries decide to enter or not enter a club. In the second stage, club members "choose" the transfer schedule designed to make outside countries abate more. In the third stage, all countries simultaneously choose their abatement levels, non-members choosing a best response.

As in the model without transfers, the only real choice for club members is in the first stage, whether or not to be in the club. If they do decide to become members then they are contractually obligated to abate and offer a transfer schedule at a level dictated by club size. As in the model without transfers, we assume abatement levels are set at q^K so as to maximize payoffs for the K members. As for the transfer level, that too can be club size dependent and we will denote it $\theta^K(\bullet)$. We will explore the equilibrium properties of different specifications of $\theta^K(\bullet)$. Note that the best response for non-members is going to be transfer dependent; call it $q^*(\theta^K)$.

Equilibrium - As before, a club size K is an equilibrium if no member wants to drop out and no non-member wants to come in. So, if a country anticipates that $K-1$ others are going to join the club, it can a) join and abate at q^K and transfer according to the schedule $\theta^K(\bullet)$ (with $N-K$ others abating at $q^*(\theta^K)$) or b) stay out and abate at $q^*(\theta^{K-1})$ (alongside $N-K$ others) and let $K-1$ countries abate at q^{K-1}. The participation IC for members requires that a) has a higher payoff than b).

The second IC is that non-members do not want to join: abating at $q^*(\theta^K)$ (along with $N-K-1$ others) and getting the associated transfer is better than abating at q^{K+1} and providing a transfer $\theta^{K+1}(\bullet)$ alongside the other K club members.

The question of interest is - do transfers increase the highest value of K satisfying both IC and, if so, by how much? The potential driver for such expansion is that transfers increase the abatement levels of non-members. That (possibly) increases the payoffs of club members (net of transfers). In that case, club membership is more attractive.

3.2 Expanded Participation With Threshold Transfers

Let us start with a simple situation where the transfer schedules are size independent but offered only if a threshold level is hit for club size.

Definition Consider a threshold size $\widetilde{K}$. A transfer policy is a *threshold* one if there is a size independent schedule $\theta(\bullet)$ such that $\theta(q)$ is paid to a non-member for every abatement level q but only if $K \geqslant \widetilde{K}$. If the club size is $K < \widetilde{K}$, no transfers are provided.

We start by specifying a particular size independent schedule $\theta(\bullet)$. We then examine what are the associated equilibrium club sizes $\widetilde{K}$.

Recall that, in the absence of transfers, the best response abatement for a non-member is q^*. Denote

$$v^* = bq^* - \frac{1}{2}cq^{*2},$$

which is the "own" payoff to a non-member from its abatement choice q^*.[9]

Lemma 2 *To induce an abatement $q \geqslant q^*$ the minimum transfer needed is*

$$\theta(q) = v^* - [bq - \frac{1}{2}cq^2], \tag{6}$$

and that is independent of club size K and of the abatement choices of other countries, whether they be members or non-members.

Proof. A non-member can always abate at q^* and receive no transfers. Hence, if other non-members are abating $\widetilde{q}$, a threshold transfer policy is incentive compatible only if,

$$b[q + Kq^K + (N - K - 1)\widetilde{q}] - \frac{1}{2}cq^2 + \theta(q) \geq v^* + b[Kq^K + (N - K - 1)\widetilde{q}].$$

There is no reason to offer anything but the least amount of transfer required. Hence, turning the above into an equality, dropping common terms and re-arranging, we get Eq. 6. The lemma is proved. ∎

From this point on, the transfer schedule given by Eq. 6 is the one we will study. The next question is how many countries have an incentive to join a club if they are required to make transfers according to that schedule. The first thing we need to do is compute the payoff - net of transfers - that a club member would get.

Note that the way we have constructed the transfer schedule $\theta(q)$ renders a recipient indifferent across abatement levels. Which of these various q is most profitable for club members to induce is hence the question to address. Club members' payoffs at that most profitable q would be what a potential signatory to membership would then consider.

Suppose we have a club of size K and it induces an abatement q. Then, invoking Eq. 5, a club member's payoff is

$$b[Kq^K + (N - K)q] - \frac{1}{2}cq^{K^2} - \frac{(N - K)}{K}\theta(q). \tag{7}$$

---

[9] In other words, the part of the payoff due to other $j's$ abatement choices, $b\sum_{j \neq i} q_j$, is omitted.

Since we are interested in finding the q that has maximum payoff for members, we need only retain terms involving q and state the optimization as

$$\max_q [bq - \frac{\theta(q)}{K}],$$

or

$$\max_{\tilde{q}} [bq + \frac{bq - \frac{1}{2}cq^2}{K}]$$

which gives the most profitable q to induce, $q(K)$ as

$$q(K) = \frac{b}{c}(K + 1). \tag{8}$$

Consequently, the payoff for a club member given abatements of q^K from fellow members and $q(K)$ from non-members is

$$\Pi^{K,\theta} = b[Kq^K + (N-K)q(K)] - \frac{1}{2}cq^{K2} - \frac{(N-K)}{K}\theta(q(K)).$$

Plugging in for $q^K (= K\frac{b}{c})$, $q(K) (= \frac{b}{c}(K+1))$ and for $\theta(q(K))$ from Eq. 6, and after doing some algebra, we get that

$$\Pi^{K,\theta} = B^\theta(N, K) - CK^2, \tag{9}$$

where $B^\theta(N, K) = \frac{b^2}{c}[\frac{1}{2}K(N+K) + N - K]$ and $C = \frac{b^2}{2c}$.

Recall that, when no transfers were offered, the payoff to a member - as computed in Eq. 2 - is $B(N, K) - CK^2$ where $B(N, K) = \frac{b^2}{c}[K^2 + N - K]$. In other words, net of transfers, a club member now has a higher payoff equal to $\frac{1}{2}KN$. This makes club membership more attractive and can, potentially, increase club size in equilibrium.

Proposition 3 *A threshold transfer strategy is an equilibrium if the threshold $\widetilde{K}$ satisfies the condition*

$$2\widetilde{K} + \frac{3}{\widetilde{K}} - 4 \leqslant N \tag{10}$$

This condition is satisfied for a range of thresholds $[k(N), K(N)]$ with $\widetilde{K} = \frac{N}{2} \in (k(N), K(N))$. $k(N)$ and $K(N)$ are the smaller and larger solutions, respectively, of $2K + \frac{3}{K} - 4 = N$. The largest club size is $K(N)$. It increases in N and $\frac{K(N)}{N}$ converges to $\frac{1}{2}$ as $N \longrightarrow \infty$.

Proof. What would make $\widetilde{K}$ an equilibrium club size? The first IC to check is that no member wants to leave the club, i.e.,

$$\Pi^{\widetilde{K},\theta} = B^\theta(N, \widetilde{K}) - C\widetilde{K}^2 \geqslant \pi^{\widetilde{K}-1,0} = B(N, \widetilde{K} - 1) - C,$$

11

where $\pi^{\widetilde{K}-1,0}$ is the payoff from exiting the club. That causes two effects: the club size falls to $\widetilde{K}-1$ and - being below the threshold - transfers drop to zero (which is denoted by the 0 in the superscript).

A little bit of algebra - using the values of $B^\theta(\bullet)$, $B(\bullet)$ and C - yields the implied condition of Eq. 10. It is straightforward to check that $\widetilde{K} = \frac{N}{2}$ satisfies Eq. 10 for all N.

The second IC that needs checking is that no non-member wants to joing the club if the size is $\widetilde{K}$, i.e., that $\pi^{\widetilde{K},\theta} \geqslant \Pi^{\widetilde{K}+1,\theta}$. It can be shown that for any $K \geqslant \widetilde{K}$,

$$\pi^{K,\theta} = b^\theta(N,K) - C,$$

where $b^\theta(N,K) = \frac{b^2}{c}[(N-1)K + N - K]$, and, recall $C = \frac{b^2}{2c}$. Using the value of $\Pi^{\widetilde{K}+1,\theta}$ as computed in Eq. 9, a bit of algebra shows that this IC holds provided

$$\widetilde{K} \geqslant \frac{N-1}{N-2},$$

and that holds since it holds whenever $K \geqslant 2$. Hence, the binding constraint on the threshold strategy is the first one given by Eq. 10. The range of possible equilibrium thresholds has hence been established.

Since the inequality defines a convex quadratic function, evidently the largest club size is associated with the larger root. Simple computation yields

$$K(N) = \frac{4 + N + \sqrt{((4+N)^2 - 24)}}{4},$$

from which it follows that

$$\frac{K(N)}{N} = \frac{1}{N} + \frac{1}{4} + \sqrt{\left(\frac{1}{16} + \frac{1}{2N}\left(1 - \frac{1}{N}\right)\right)},$$

which converges to $\frac{1}{2}$. The proof is complete. ∎

Remark - Note that a club member's payoff, as shown in Eq. 9 is given by $B^\theta(N,K) - CK^2$ where $B^\theta(N,K) = \frac{b^2}{c}[\frac{1}{2}K(N+K) + N - K]$ and $C = \frac{b^2}{2c}$. By contrast, a non-member's payoff is given by $b^\theta(N,K) - C$, where $b^\theta(N,K) = \frac{b^2}{c}[(N-1)K + N - K]$, as seen directly above. It is easy to check that a non-member has a higher payoff than a club member since they are not committed to high abatement levels as club members are.

3.2.1 Generalized Threshold Transfers

The threshold transfer strategy can be generalized to the following off-path rule.

Definition (Off-path Transfers) If in a club of size $\widetilde{K}$ one signatory drops, then the remaining $\widetilde{K}-1$ signatories commit to an abatement schedule of $q^{\widetilde{K}-1} = \frac{(\widetilde{K}-1)b}{c}$ and a transfer that makes non-signatories abate $\tilde{q}^{\widetilde{K}-1} = \frac{b}{c}x$ where $1 \leq x \leq \widetilde{K}$.

The idea here is that there need not be a complete drop in transfers made to non-club countries if some club country deviates. Club countries, as we shall

see, can decide a drop in transfers that would be sufficient to deter deviation by club countries.

Proposition 4 *There is an equilibrium with $\widetilde{K} \geq \frac{N}{2}$ when club members provide transfers and the off-path rule followed is given in the Definition (Off-path Transfers). The pair of highest off-path abatement quantity x and club size $\widetilde{K}$ that can be sustained in equilibrium is given by*

$$x = \frac{N + \widetilde{K} - \widetilde{K}^2 - \frac{3}{2} + \frac{N\widetilde{K}}{2}}{N - \widetilde{K}}$$

Proof. Similar to Proposition 3, we need to check two IC conditions. Note that the condition ensuring that the club of size $\widetilde{K}$ will not expand to size $\widetilde{K} + 1$ is the same as before,

$$\widetilde{K} \geqslant \frac{N - 1}{N - 2},$$

and is true whenever $K \geq 2$. Therefore, our task is to ensure the other incentive compatibility condition, namely that no club country would want to exit the club, now with the above-mentioned rule in Definition ((Off-path Transfers). The IC condition for no member wanting to leave the club,

$$\Pi^{\widetilde{K},\theta} = B^\theta(N, \widetilde{K}) - C\widetilde{K}^2 \geqslant \pi^{\widetilde{K}-1,x} = b[(\widetilde{K} - 1)^2 + 1 + (N - \widetilde{K})\frac{xb}{c}] - C$$

where $\pi^{\widetilde{K}-1,x}$ is the payoff from exiting the club. In this case, there are two effects, the club size falls to $\widetilde{K} - 1$, and the transfers drop to an amount that ensures $x\frac{b}{c}$ is abated by every country outside the club. A little bit of algebra shows that the maximum sustainable abatement amount for non-club countries is:

$$x = \frac{N + \widetilde{K} - \widetilde{K}^2 - \frac{3}{2} + \frac{N\widetilde{K}}{2}}{N - \widetilde{K}}$$

It is easy to check that $\widetilde{K} = \frac{N}{2}$ is always an equilibrium, and that the highest $\widetilde{K} > \frac{N}{2}$. ∎

The above result can be seen from Figure 1 where the club size is plotted on the x-axis and the required off-path transfer is on the y-axis. $N = 150$, which is close to the number of countries represented in international bodies.

The above result can be seen from Figure 1 where the club size K is plotted on the x-axis and the required off-path transfer x is on the y-axis for a model with $N = 150$, which is close to the number of countries represented in international bodies. The figure indicates that the maximum achievable club size is roughly $77 > N/2$.

Remark - Note that the earlier observation, that non-members have a higher payoff in equilibrium compared to club members, continues to be true. This is because that is not an off-path comparison but rather one for any K that is on-path. Generalized threshold strategies can change the value of K but not the on path ranking of payoffs.

3.3 Pareto Improvement With Transfers

In this subsection we look at two questions: how do equilbrium payoffs change with club size? And, how does the transfer model payoffs compare with the no transfer ones.

Note first that a larger club is good for both club countries and non-club countries, as a look at their payoffs, given by $\Pi^{\widetilde{K},\theta}$ and $\pi^{\widetilde{K},\theta}$ respectively, indicates. The below result directly follows and we state it without proof.

Proposition 5 *Equilibrium payoffs increase in $\widetilde{K}$ for both signatories and non-signatories*

Second, we note that transfers do indeed cause a Pareto improvement, increasing payoffs for both club and non-club members. This emphasizes the importance of transfers in achieving larger club sizes.

Proposition 6 *Transfers effect a Pareto improvement in equilibrium for both club members and non-members.*

Proof. For any club with threshold size $\widetilde{K}$, the difference in payoff for a club country between a world with transfers and a world without transfers is

$$\Pi^{\widetilde{K},\theta} - \Pi^{\widetilde{K}} = (N - \widetilde{K})\frac{\widetilde{K}}{2} > 0$$

For each non-club country, the difference in payoffs with transfers and without transfers is:
$$\pi^{\widetilde{K},\theta} - \pi^{\widetilde{K}} = (N - \widetilde{K} - 1)\widetilde{K} > 0$$

Hence, for both club and non-club members, transfers lead to a Pareto improvement. ∎

3.4 Heterogeneity

The results in the previous section leads one to ask the following question: which countries should be in the club? This of course requires some concept of heterogeneity since in the perfectly homogeneous model every country is the same. Consider then, that each country i differs in the cost of its abatement c_i. We think of this cost as the labor cost of abatement. Or, equivalently, the cost it takes for a country to create investments such as forests, that are important carbon sinks.

This simple source of heterogeneity allows us to order the countries that will join the club. This result is applicable for the generalized threshold transfer system, and a fortiori, applies to the simple threshold strategy as well. It provides a rationale for some countries for being the club while others staying outside.

Proposition 7 *If the c_i differ, then greatest incentive to join is for the country with the highest c_i. Hence, the largest club size K is derived when the K countries with the highest c_i join the club.*

14

Proof. In the final stage, let there be a club with threshold size $\widetilde{K}$. The club countries will set a target for each non-club member that they will try to achieve from transfers. This target is obtained by solving for each non-club country i:

$$\max_{\tilde{q}} b\tilde{q} + \frac{b\tilde{q} - \frac{1}{2}c_i\tilde{q}^2}{\widetilde{K}}$$

which gives us, as before

$$q(\widetilde{K}) = \frac{b}{c_i}(\widetilde{K} + 1) \tag{11}$$

Therefore, each non club country will be induced to abate $q(\widetilde{K})$. The difference now is that the abatement quantity differs by the cost of abatement for each non-club country.

As usual, each club country j will abate:

$$q(\widetilde{K}) = \frac{\widetilde{K}b}{c_j}$$

As before, the punishment off-path will be to induce $\frac{xb}{c_i}$ abatement by the non-club countries if the size goes below the threshold size, to $\widetilde{K} - 1$. Assume that the $\widetilde{K}$'th country deviates and leaves. The upper-bound on the value of x can be calculated, taking care now of the different costs of abatement for each country:

$$\Pi_{\widetilde{K}}^{\widetilde{K},\theta} \geq \pi_{\widetilde{K}}^{\widetilde{K}-1,x}$$

The subscript $\widetilde{K}$ accounts for the payoff of the $\widetilde{K}$'th country incorporating the heterogeneity in costs. Furthermore, the payoffs on either side can be written as,

$$B_{\widetilde{K}}^{\theta}(N, \widetilde{K}) - C_{\widetilde{K}}\widetilde{K}^2 \geq b^2[(\widetilde{K} - 1)\sum_{j=1}^{\widetilde{K}-1}\frac{1}{c_j} + \frac{1}{c_{\widetilde{K}}}\sum_{i=\widetilde{K}+1}^{N}\frac{x}{c_i}] - C_{\widetilde{K}}$$

where $B_{\widetilde{K}}^{\theta}(N, \widetilde{K}) = b^2[\widetilde{K}\sum_{k=1}^{N}\frac{1}{c_k} - \sum_{i=\widetilde{K}+1}^{N}\frac{\widetilde{K}/2-1}{c_i}]$ is the version of $B^{\theta}(N, \widetilde{K})$ under heterogeneity taking into account the different costs of abatement for each country. Similarly, $C_{\widetilde{K}} = \frac{b^2}{2c_{\widetilde{K}}}$.

Some algebra allows us to calculate the largest x that can be sustained in equilibrium

$$x \leq \frac{\sum_{j=1}^{\widetilde{K}}\frac{1}{c_j} + \sum_{i=\widetilde{K}+1}^{N}\frac{\widetilde{K}/2+1}{c_i} + \frac{\widetilde{K}}{c_{\widetilde{K}}} - \frac{\widetilde{K}^2}{2c_{\widetilde{K}}} - \frac{3}{2c_{\widetilde{K}}}}{\sum_{i=\widetilde{K}+1}^{N}\frac{1}{c_i}}$$

It is easy to see that this expression under homogeneity will give the same expression for x as in the previous subsection.

Now that we have obtained a sense of off-path transfers that sustain equilibrium, we can now consider the participation decision in the first stage. Let's

15

say some country j considers the benefit of deviating and not joining the club. First, the payoff from being in the club for any country is given by,

$$b^2[\frac{\widetilde{K}}{c_1} + ... + \frac{\widetilde{K}}{c_{\widetilde{K}}} + \frac{\widetilde{K}+1}{c_{\widetilde{K}+1}} + ... + \frac{\widetilde{K}+1}{c_N}] - \frac{1}{2}\frac{b^2}{c_j}\widetilde{K}^2 - \tilde{\theta}_j$$

The transfer $\tilde{\theta}_j$ can be obtained by observing that it will be j's fraction of total transfers made to all non-club countries. In other words, $\tilde{\theta}_j = \frac{1}{\widetilde{K}}\sum_{i=\widetilde{K}+1}^{N}\tilde{\theta}_i$. The transfer received by a country i is,

$$\tilde{\theta}_i = \frac{1}{2}\frac{b^2}{c_i}[1 - [(\widetilde{K}+1) - \frac{1}{2}(\widetilde{K}+1)^2]] = \frac{\widetilde{K}^2}{2}\frac{b^2}{c_i}$$

Hence, we can substitute in the value of $\tilde{\theta}_j$ to get the payoff from staying in the club,

$$b^2[\frac{\widetilde{K}}{c_1} + ... + \frac{\widetilde{K}}{c_{\widetilde{K}}} + \frac{\widetilde{K}+1}{c_{\widetilde{K}+1}} + ... + \frac{\widetilde{K}+1}{c_N}] - \frac{1}{2}\frac{b^2}{c_j}\widetilde{K}^2 - \frac{b^2\widetilde{K}}{2}[\frac{1}{c_{\widetilde{K}+1}} + ... + \frac{1}{c_N}]$$

The first and third terms are the same for every club country. The only thing that differentially affects the payoff from being in the club is the second term which increases as the cost c_j increases. In other words, a country with higher c_j has a greater payoff from being in the club.

The other payoff we need to calculate is that from choosing to be outside the club, noting that every non-club country is induced to abate $\frac{xb}{c_i}$, while the deviating country j obtains a payoff from its own abatement action which is exactly equivalent to complete free-riding,

$$b^2[\frac{\widetilde{K}}{c_1} + ... + \frac{\widetilde{K}}{c_{j-1}} + \frac{\widetilde{K}}{c_{j+1}} + ... + \frac{x}{c_{\widetilde{K}+1}} + ... + \frac{x}{c_N}] + \frac{1}{2}\frac{b^2}{c_j}$$

This payoff clearly decreases as c_j increases. In other words, a country with higher c_j has a lower payoff from deviating and leaving the club. This completes our proof. Naturally, the configuration of countries that leads to the biggest club size will be that when the K countries with the highest c's join the club ∎

4 Multi-Lateral Transfers and Efficiency

4.1 Model

So far we only allow club members to make transfers. This section outlines a model in which non-members can also do so. We retain the perspective that a club acts like a synthetic player, choosing its actions to optimize for the whole group. Hence, as before, the club picks transfers to best motivate non-members and picks its abatement level at a best response. Now that non-members can

also choose transfers, the club's best response abatement will not only depend on club size but also the size of the transfers from non-members.

Timing - There are three stages. In the first stage, a set of countries decide to enter or not enter a club. In the second stage, club members and non-members choose transfer schedules designed to make other countries abate more. In the third stage, all countries simultaneously choose their best response abatement levels. In stages two and three, each non-member chooses a best response for itself and the club does so as well (but as a single synthetic player).

Let K denote a club - as also the size of the club. Let q_K denote the per capita abatement of each club member and let $\theta_{Ki}(\bullet)$ denote the transfer schedule offered to a non-member i by the club. Similarly, $\theta_{iK}(\bullet)$ denotes the transfer made by a non-member to the club. We continue to assume transferable utility; a non-member's payoff is

$$\pi_i(q;\theta) = b\sum_{k\in N} q_k - \frac{1}{2}cq_i^2 + \sum_{j\neq i,K} \theta_{ji}(q_i) + \theta_{Ki}(q_i) - \sum_{j\neq i,K}\theta_{ij}(q_j) - \theta_{iK}(q_K), \ (12)$$

where j is another non-member. Similarly, the payoff of a club member is

$$\Pi_i(q;\theta) = b\sum_{k\in N} q_k - \frac{1}{2}cq_K^2 + \frac{1}{K}\sum_{i\neq K}\theta_{iK}(q_K) - \frac{1}{K}\sum_{i\neq K}\theta_{Ki}(q_i). \qquad (13)$$

Equilibrium - In the third stage, q_K and q_i are chosen to, respectively, maximize Eqs. 13 and 12. In stage two, keeping that in mind, θ_K and θ_i are chosen so as to maximize payoffs of each player and the club. That determines a value in the game for a non-member as well as a club member and these values may vary depnding on the size of K. In stage one, a club size K is an equilibrium if no member wants to drop out and no non-member wants to come in.

The question of interest is - does allowing global transfers, and not restricting them to only come from the club, increase the highest value of K in equilibrium and, if so, by how much? Does it also lead to greater cumulative abatement?

4.2 The Common Agent Set-up

Suppose we are at the second stage, i.e., club size K has already been determined. Without loss, suppose we label countries such that $i = 1, ... N - K$ are the non-members.

What remains to be determined are the transfer choices in stage two and the consequent abatements in stage three. We first show that our setup is a special case of Bernheim-Whinston (1986). We then characterize the precise equilibria in our setting.

In stage two, each non-member country i chooses transfers θ_{ij} and θ_{iK} where $\theta_{ij} \geq 0$, $\theta_{iK} \geq 0$; $i, j = 1, ..., N - K$, $i \neq j$. Similarly, the club chooses θ_{Ki}, $i = 1, ..., N - K$. These choices incentivize the abatement choices in stage three.

Consider a non-member i. In stage three, it picks q_i to maximize Eq. 12. Dropping all terms that do not involve q_i gives us the following optimization:

$$\max_{q_i}[bq_i - \frac{1}{2}cq_i^2 + \sum_{j \neq i,K} \theta_{ji}(q_i) + \theta_{Ki}(q_i)],$$

which leads to a solution $q_i(\theta_{ji}, \theta_{Ki}, j \neq i, K)$. Note that this optimization is independent of the contemporaneous choices q_j and q_K. It is also independent of transfers targeted at countries other than i. Put differently, starting at stage two, we have a Bernheim-Whinston (1986) Common Agency problem with non-member i a common agent for the $N - K - 1$ Principals who are the other non-members and, additionally, the club also acting as a Principal. By Bernheim-Whinston (1986) we know that there is at least one equilibrium, starting in stage two, that involves i playing the efficient action, i.e., $\widehat{q} = \arg\max_{q_i}[Nbq_i - \frac{1}{2}cq_i^2]$.

We can - alternatively - construct that equilibrium directly in our model.

Lemma 8 *For any club size K, there is a continuation equilibrium in the sub-game in which each country, whether they be a non-member or a club member, abates at $\widehat{q}$, for all countries $i = 1, ..., N - K$ and club K.*

Proof. For every other non-member $j \neq i$, consider the following transfer scheme:

$$\theta_{ji}(\widehat{q}) = \frac{1}{N-1}\{v^* - [b\widehat{q} - \frac{1}{2}c\widehat{q}^2]\},$$
$$= 0, q_i \neq \widehat{q},$$

which is a Dirac version of the transfer scheme of the previous section given by Eq. 6 and where, recall, $v^* = bq^* - \frac{1}{2}cq^{*2}$. The club offers an aggregated version of that transfer,

$$\theta_{Ki}(\widehat{q}) = \frac{K}{N-1}\{v^* - [b\widehat{q} - \frac{1}{2}c\widehat{q}^2]\},$$
$$= 0, q_i \neq \widehat{q}.$$

We first show that these are best responses to each other. Consider the best response transfer choice of a non-member j. If it provides the stated transfer at $\widehat{q}$, then country i would abate at $\widehat{q}$ and $j's$ payoff (due to that abatement alone) would be

$$b\widehat{q} - \theta_{ji}(\widehat{q}) = \frac{Nb\widehat{q} - \frac{1}{2}c\widehat{q}^2}{N-1} - \frac{v^*}{N-1}. \tag{14}$$

It can also give less than the stated amount in which case country i would pick q^* and $j's$ consequent payoff would be

$$bq^*.$$

The transfer induced payoff, Eq. 14 is better iff

$$Nb\widehat{q} - \frac{1}{2}c\widehat{q}^2 \geqslant (N-1)bq^* + v^* = Nbq^* - \frac{1}{2}cq^{*2}, \tag{15}$$

18

and that holds by definition. Consider instead the best response of the club. If it provides the stated transfer at $\widehat{q}$, then country i would abate at $\widehat{q}$ and club $K's$ payoff (due to that abatement alone) would be

$$Kb\widehat{q} - \theta_{Ki}(\widehat{q}) = K[\frac{Nb\widehat{q} - \frac{1}{2}c\widehat{q}^2}{N-1} - \frac{v^*}{N-1}]. \tag{16}$$

It can also give less than the stated amount in which case country i would pick q^* and $K's$ consequent payoff would be

$$Kbq^*. \tag{17}$$

The transfer induced payoff, Eq. 14 is better iff the RHS of Eq. 16 is greater than Eq. 17 and that, of course, is implied by Eq. 15. Now consider the club.

Now consider the club. If no transfers are provided they would pick $q^K = \arg\max_q[bKq - \frac{1}{2}cq^2]$. Call the associated value v^K. Evidently, the club is indifferent between q^K and $\widehat{q}$ provided the transfer $\theta_{iK}(\widehat{q})$ each non-member makes to the club satisfies

$$Kv^K = K[bK\widehat{q} - \frac{1}{2}c\widehat{q}^2 + \frac{N-K}{K}\theta_{iK}(\widehat{q})] \tag{18}$$

where Eq. 18 has the appropriate adjustments to accounte for the fact that there are $N - K$ non-members and the club size is K. Suppose, as before, that transfers are only made for $\widehat{q}$. A non-member is willing to do that provided payoffs net of transfers are better than the no transfer outcome, i.e.,

$$bK\widehat{q} - \theta_{iK}(\widehat{q}) \geqslant bKq^K.$$

Substituting from Eq. 18, that turns out to be equivalent to

$$Nb\widehat{q} - \frac{1}{2}c\widehat{q}^2 \geqslant (N-K)bq^K + v^K = Nbq^K - \frac{1}{2}cq^{K^2}$$

and that holds by definition. The lemma is proved. ∎

Now we go to the first stage. Regardless of how many countries sign on to be in the club, by the Lemma above, the eventual abatement is the utilitarian optimal $\widehat{q}$ and the transfers all cancel out. So, every country's payoff, independent of K, is $Nb\widehat{q} - \frac{1}{2}c\widehat{q}^2$. Hence, one possible equilibrium in the participation stage is for each country to sign on, i.e., for $K = N$. Hence, we have proved

Proposition 9 *There is an equilibrium with $K = N$ and every country abating at the UPO $\widehat{q}$.*

5 Sequential Transfers and Unique Efficiency

Whilst efficient abatement is an equilibrium in the club model studied in the last section, it is just one of many possible equilibria. That leaves open the question of whether there is a club model in which efficient abatement is the

only equilibrium. That is exactly what we do in this section. We do that by motivating club formation through a *sequential* design. In this design, the transfer schedules are chosen sequentially. The club sets things off by announcing how much it will pay each of the non-members to encourage abatement by them and then non-club members sequentially decide how much, if anything, to pay other countries including the club.

Timing - There is, again, a first stage in which countries choose whether or not to be in the club. This is followed by a second stage with $N - K + 1$ sub-stages. As above, and without loss, suppose non-members are labeled $1, ...N - K$. In sub-stage 0, the club announces its commitment to non-members $\theta_{K,j}(\bullet)$, $j = 1, ...N - K$. Call the vector θ_K. The commitments can be zero. In sub-stage 1, (non-member) country $N - K$ picks transfers for the club $\theta_{N-K,K}(\bullet)$ and for other non-members $\theta_{N-K,j}(\bullet)$, $j = 1, ...N - K - 1$. Call the vector θ_1. Then, in sub-stage 2, (non-member) country $N - K - 1$ picks transfers for the club and for other non-members. And so on, till non-member country 1. Then there is a final stage in which every country (simultaneously) chooses their best response abatement. As before, for a club, the choices in stages two and three are collectively made as if the club is a single (synthetic) player.

Payoffs incorporate transferable utility and are given by Eqs. 12 and 13.

Equilibrium - In the third stage, q_K and q_i are chosen to, respectively, maximize Eqs. 12 and 13. In every sub-stage of stage two, θ_K and then, sequentially, non-member commitments θ_i are chosen looking ahead at subsequent choices. For instance, when country 1 chooses θ_1 it already knows the club's choice θ_K as well as θ_i, $i \neq 1$. Call that vector of vectors θ_{-1}. Hence, it chooses a best response θ_1 given θ_{-1} thereby creating a best response function $\theta_1(\theta_{-1})$.

Look now at the best response choice of the penultimate chooser, country 2. At that point, it already knows the club's choice θ_K as well as θ_i, $i \neq 1, 2$. Call that vector of vectors θ_{-12}. In making its choice, 2 takes account of the subsequent best response function $\theta_1(\theta_{-1})$ as well as θ_{-12}. Hence, it chooses a best response function $\theta_2(\theta_{-12})$. And so on.

Given a sequence, these sequential transfer choices determine a value in the game for every non-member as well as for a club member. In stage one, using these values, a club size K is determined as an equilibrium such that no member wants to drop out and no non-member wants to come in.

The question of interest is - do sequential transfers narrow the range of possible equilibrium abatements while retaining efficiency as an option?

5.1   The Sequential Common Agent Set-up

Suppose we are at the second stage, i.e., club size K has already been determined. What remains to be determined are the sequential transfer choices in stage two and the consequent abatements in stage three. We first show that our setup is a special case of Prat-Rustichini (2003) and Dutta-Siconolfi (2024). Their results conceptually pin down the equilibrium possibility in stages two and three of our game. We discuss their results and then directly and fully characterize equilibria in our setting.

Consider a non-member i's choice in stage three. As in the rest of the analysis, it picks q_i to maximize Eq. 12 and that optimization is independent of the contemporaneous choices q_j and q_K and is also independent of transfers targeted at players other than i. Hence, i is, as in the previous section, a common agent for the $N - K - 1$ Principals who are the other non-members and, additionally, the club also acting as a Principal. Unlike Bernheim-Whinston (1986) though, the Principals choose transfers sequentially and hence their result cannot be directly applied.

Prat-Rustichini (2003) amended the Bernheim-Whinston (1986) setting to allow sequential transfers in the common agent setting, allowing the Principals to pick compensation functions for the agent sequentially. In turn, their problem is a special case of Dutta-Siconolfi (2024) that considers the play of any stage-game - not just a Common Agent game - and models a pre-play phase in which the players sequentially choose side-payments for each other. In the Common Agent setting, Prat-Rustichini (2003) show that there there is a unique action taken in equilibrium by the agent and that is the socially optimal one. Dutta-Siconolfi (2024) show that result holds more generally in any game. In our setting, the two results assert that equilibrium action is unique and efficient. However, the results do not tell us what the on-path transfer schedules will be, just that their exist schedules that support subsequent efficient abatement.

To exactly characterize the transfers, we directly construct them.

Lemma 10 *For any club size K, there is a unique continuation equilibrium action vector and that is one in which each country, whether they be a non-member or a club member, abates at $\widehat{q}$, and this is true for all countries $i = 1, ..., N$. On path, transfers are minimally sufficient to induce that abatement level.*

Proof. Suppose a non-member i is the common agent.

Consider the last transfer chooser, country 1 and consider its choice of θ_{1i} chosen to influence the abatement choice of country i. At that point all commitments except those of country 1 have already been announced; denote that vector θ_{-1}. Denote further by $v_i(\theta_{-1})$ the highest payoff i can obtain given those commitments, i.e.,

$$v_i(\theta_{-1}) = max_{q_i}\{bq_i - \frac{1}{2}cq_i^2 + \sum_{j=2}^{N-K} \theta_{ji}(q_i) + \theta_{Ki}(q_i)\}$$

Clearly, i can do no worse than get a payoff of $v_i(\theta_{-1})$ no matter what country $1's$ commitment turns out to be since such a commitment can only increase that value. However, 1 can induce any abatement q provided it gives a transfer $\theta_{1i}(q)$ that satisfies

$$bq - \frac{1}{2}cq^2 + \sum_{j=2}^{N-K} \theta_{ji}(q) + \theta_{Ki}(q) + \theta_{1i}(q) = v_i(\theta_{-1}). \qquad (19)$$

The best such q from the perspective of country 1 can be deduced by maximing its payoffs on account of country $i's$ abatement, i.e., by solving

$$
\max_q \{bq - \theta_{1i}(q)\}
$$

$$
= \max_q \{2bq - \frac{1}{2}cq^2 + \sum_{j=2}^{N-K} \theta_{ji}(q) + \theta_{Ki}(q)\} - v_i(\theta_{-1})
$$

$$
\equiv v_{1i}(\theta_{-1}),
$$

where the second expression follows by substituting from Eq. 19 and $v_{1i}(\theta_{-1})$ is the highest value that country 1 can derive from the common agent $i's$ abatement given the commitments of all donors but 1. Between them, countries 1 and i hence maximally get a value we denote $V_{1i}(\theta_{-1})$:

$$
V_{1i}(\theta_{-1}) \equiv v_{1i}(\theta_{-1}) + v_i(\theta_{-1}) = \max_q \{2bq - \frac{1}{2}cq^2 + \sum_{j=2}^{N-K} \theta_{ji}(q) + \theta_{Ki}(q)\}. \quad (20)
$$

Consider now the second to last transfer chooser, country 2 and consider its choice of θ_{2i} again chosen to influence the abatement choice of country i. At that point all commitments except those of countries 1 and 2 have already been announced; denote that vector θ_{-12}.

We can deduce from the immediately preceding that the pair of countries 1 and i can do no worse than get a payoff of $V_{1i}(\theta_{-12}, \theta_2 = 0)$ no matter what country $2's$ commitment is. However, 2 can induce any abatement q provided it gives a transfer $\theta_{2i}(q)$ that satisfies

$$
2bq - \frac{1}{2}cq^2 + \sum_{j=3}^{N-K} \theta_{ji}(q) + \theta_{Ki}(q) + \theta_{2i}(q) = V_{1i}(\theta_{-12}, \theta_2 = 0). \quad (21)
$$

The best such q from the perspective of country 2 can be deduced by maximing its own payoffs on account of country $i's$ abatement, i.e., by solving

$$
\max_q \{bq - \theta_{2i}(q)\}
$$

$$
= \max_q \{3bq - \frac{1}{2}cq^2 + \sum_{j=3}^{N-K} \theta_{ji}(q) + \theta_{Ki}(q)\} - V_{1i}(\theta_{-12}, \theta_2 = 0)
$$

$$
\equiv v_{2i}(\theta_{-12}),
$$

where the second expression follows by substituting from Eq. 21 and $v_{2i}(\theta_{-1})$ is the highest value that country 2 can derive from the common agent $i's$ abatement given the commitments of all donors but 1 and 2. Between them, countries $1, 2$ and i hence get a value we denote $V_{12i}(\theta_{-12})$:

$$
V_{12i}(\theta_{-12}) \equiv v_{2i}(\theta_{-12}) + V_{1i}(\theta_{-12}, \theta_2 = 0) = \max_q \{3bq - \frac{1}{2}cq^2 + \sum_{j=3}^{N-K} \theta_{ji}(q) + \theta_{Ki}(q)\}.
$$

The logic of the above arguments applied recursively to all non-members says that collectively, given any club commitment θ_{Ki}, they get a value we denote $V_{-K}(\theta_{-K})$:

$$V_{-K}(\theta_{-K}) = \max_q \{(N-K)bq - \frac{1}{2}cq^2 + \theta_{Ki}(q)\}.$$

The club can induce any abatement q provided it gives a transfer $\theta_{Ki}(q)$ that satisfies

$$(N-K)bq - \frac{1}{2}cq^2 + \theta_{Ki}(q) = V_{-K}(\theta_{-K} = 0). \qquad (22)$$

The best such q from the perspective of the club can be deduced by maximing its club payoffs on account of country $i's$ abatement, i.e., by solving

$$\max_q \{Kbq - \theta_{Ki}(q)\}$$
$$= \max_q \{Nbq - \frac{1}{2}cq^2\} - V_{-K}(\theta_{-K} = 0)$$

where the second expression follows by substituting from Eq. 22. The solution is, of course the UPO abatement $\widehat{q}$.

The same logic works if the abatement chooser is the club and the non-members sequentially incentivize it to pick their preferred abatement. We have hence shown that the unique equilibrium abatement in stage three is the optimal one - and this is true for every country. We have also exactly - though implicitly - characterized the on path transfers in stage two. The lemma is proved. ∎

Now we look at the very first stage, that of choosing whether or not to be in the club. The incentives now will depend on the labeling of countries $1, ...N$ since that will determine their place in the sequence if they choose not to enter the club. Depending on the parameters, it might be more profitable to be outside the club if the label number is higher - a last mover advantage - or lower - a first mover advantage. In the symmetric model we are analyzing that will in turn determine who - and how many - join the club. Regardless though, due to the lemma above, the abatement outcome will be efficient. In other words we have the following result:

Proposition 11 *There is an equilibrium club size $K \geqslant 1$. No matter how large the club is, the unique subgame perfect equilibrium is one in which the abatement is Utilitarian Pareto Optimal for all countries, i.e., $q_i = \widehat{q}$, $i = 1, ...N$.*

6 Literature and Extensions

6.1 Literature

The first important reference point for our work is Barrett (2003) which offers a textbook treatment of international environmental treaties. Chapter 7 of Barrett (2003) models the treaty participation game. It starts with a linear model

of benefits and costs with three stages, a participation stage where countries decide to be in the treaty, then stage two where the treaty countries cooperatively decide their abatement, and the last stage where non-treaty countries decide their abatement. The general result that emerges is that large treaty sizes are not attainable unless the parametrization implies very little benefit of forming large treaties. The chapter considers different specifications of benefits and costs, including of a linear-quadratic, concluding that the result that large effective treaties are not attainable in equilibrium is robust.

Chapter 13 of Barrett (2003) then considers the idea of side-payments. The result achieved is that side-payments are only effective (i) if the countries are "strongly" asymmetric, and (ii) they are accompanied by the threat of reduction in abatement by club countries.

We maintain a similar structure as in Barrett (2003, 2005) in two of our three stages - the first, when countries decide whether to participate in the club, and the third, when club members cooperatively choose abatement. Where we differ is in the second transfer stage. In the first half of our paper, we model bilateral transfers alone - where club countries give transfers to non-club countries and in the second half we study multi-lateral transfers that can go in both directions. In either case, and this is one key difference with Barrett (2003), our transfers incentivize abatement directly. By contrast, his transfers aim to incentivize non-club countries to join the club. And that makes a significant difference to the results.

Furthermore, unlike Barrett (2003), we do not require any asymmetry for our results, although heterogeneity may help by expanding the club and by providing order to the countries that become part of the club. Nor do we require club countries to reduce their own abatement if the club size falls.

The idea of treaty participants acting collectively has been modeled by several papers in the literature, see Hoel (1992), Carrao and Siniscalco (1993), Barrett (2005). In particular, Carrao and Siniscalco (1993), in a modelling environment similar to the standard treaty set-up, claim that side-payments cannot increase the size of the treaty. Their impossibility result, however, depends on added constraints that are extraneous to the model. First, they only consider transfers to non-treaty countries that are lump sum. Second, the transfer recipient country is required to abate at the same level as the donor countries. We show that (i) having a transfer schedule (rather than lump sum payments), and (ii) inducing recipient countries to abate $K + 1$ while club countries abate K allows for a Pareto improvement in abatement, and larger club sizes. The impossibility result does not hold without those additional constraints.

In addition, to achieve positive results, Carrao and Siniscalco (1993) introduce different forms of commitment, which essentially prohibits an initial set of treaty countries from leaving the treaty. With this commitment, side-payments can increase the size of the treaty. Our paper does not require any such form of commitment. Countries choose their abatement and transfers willingly, and club participation involves incentive compatibility. The transfers are aimed at incentivizing non-club countries to abate more. This is an important reason why our model is closely related to Nordhaus' idea of climate clubs, which is a group

of countries that induce outside countries to abate more, as we demonstrate in the first part of our paper.

Another relevant reference point are constituted by the papers by Chander and Tulkens (1995, 1997). They obtain a result of full participation that is based on the idea that if any country deviates and leaves the treaty, then the remaining treaty members leave the treaty and act non-cooperatively. The beliefs needed to sustain this equilibrium ensures that no country or coalition of countries can benefit by a deviation. We do not require the need for beliefs that sustain or drive the abatement decision by the club countries. The countries always choose their abatement level based on their decision to be in the club or not. In other words, if any country deviates, the other countries continue abating as a group. We also do not depend on the idea of the core in our paper, that underlies these series of papers.

The economics literature on climate clubs, a idea proposed by Nordhaus (2015), is relatively newer and therefore smaller. Nordhaus' proposal centers on the importance of trade sanctions that induce non-club participants to abate more. Nordhaus (2021) provides a modification of this approach where they combine trade sanctions with technological advancement in a multi-period model, and provide numerical results to support the importance of both benefits. This latter idea is related to Heal (1992) who considers the cost reduction achieved when more countries adopt higher abatement, thereby encouraging more participation.

Another strand of the literature looks at dynamic climate agreements. A widely used framework is Dutta and Radner (2004, 2009), that has a "static reduction," and which we use in Section 6.2 when considering a dynamic model. This framework has been used in several papers. Notable among them is Battaglini and Harstad (2016) look at the interplay of coalition size, the completeness of contracting, and contracting duration. Harstad (2023) analyzes the idea of countries making pledges which then have to be unanimously approved.

Lastly, there is a literature on side payments under the concept of common agency, a literature started by Bernheim and Whinston (1986). Under conditions provided by Bernheim-Whinston (1986), an equilibrium with full participation is implied in our set-up. We contribute by characterizing the full set of equilibria that can be achieved. Finally, our sequential equilibrium set-up builds on Prat-Rustichini (2003) and Dutta-Siconolfi (2024) that allows us to characterize the unique SPE that is achieved in Section 5.

6.2   Extensions

There are three directions in which the model can be readily extended.

Heterogenity - As we saw in Section 3.4, we can make the cost of abatement heterogenous and still derive a prediction on which countries are more likely to be in a climate club. It is our belief that one can do something similar by making the public good benefit parameter heterogenous instead. Whether we will still be able to derive a clear-cut result if both cost and benefit paramters are heterogenous, remains to be explored.

When it comes to the multi-lateral models of Sections 4 and 5, absolutely nothing changes in the results if we introduce heterogenity in both parameters. This is because the analysis essentially depended on isolating each country as a common agent and then focusing transfers by the other countries on that common agent. Since no comparisons need be made across countries, the exact same analysis - and hence results - would work in the heterogenous case.

Separable Payoffs - What if payoffs do not have the linear quadratic form studied in this paper? Turns out what is critical is separability.

Let us take that in two steps. Suppose that benefits and costs to abatement remain separable. That seems like a natural assumption since benefits depend on the public good nature of collective abatement while costs depend on the size of private abatement. In that setting, absolutely nothing changes if costs are not quadratic (but benefits remain linear). Of course, the exact number of treaty signatories in the models of Sections 2 and 3 are not possible to compute. However, the qualitative conclusions that club size is small without transfers and expanded by adding transfers, remains unchanged. Moreover, the common agent reduction of Sections 4 and 5 are wholly unchanged and so are the results. Again, keeping in mind that the exact formula for the Utilitarian Pareto Optimum abatement cannot be exactly computed if costs are, say, any convex function, rather than being exactly quadratic.

If the benefit function is not linear, then matters can get a little more complicated. If they remain separable, the analysis still applies. For example, if the total benefit is the sum of each benefit possibly filtered through a functional form - say, $b \sum_{k \in N} q_k^{\alpha}$, for some α in $(0, 1]$ - then there is no problem and we can readily apply the methodology of this paper. In Sections 2 and 3, we could even compute club size given a specific functional form for the private costs. In Sections 4 and 5, we can apply the common agent filter.

If the benefit function is not separable, say if it is $b(\sum_{k \in N} q_k)^{\alpha}$, for some α in $(0, 1]$, then the analysis would be more complicated. Our conjecture is that - even though exact club sizes would be difficult to compute - the results would still hold. This is partly informed by the knowledge that the results of Section 5 - the sequential multi-lateral model - would definitely hold. This is because the more general treatment of the problem in Dutta-Siconolfi (2024) considers any game, not just those with separable payoffs.

Dynamic Model - One clear shortcoming of the analysis is that the model is static.[10] There is one widely studied dynamic climate model that does fit into the framework here. This is the model introduced in Dutta-Radner (2004, 2009) which has been used by, among others, Harstad (2016, 2023), Chander (2017) and Battaglini and Harstad (2016), since it has a "static reduction". In that model, the stock of greenhouse gases (GHGs) g_t builds up between t and

---

[10] Of course, the same criticism can be levied against the vast literature that is summarized in Barrett (2004).

$t + 1$ through the transition equation

$$g_{t+1} = \sigma g_t + \sum_{i=1}^{N} e_{it}, \tag{23}$$

where $1 - \sigma \in (0, 1))$ is the "depreciation" (photosynthesis) factor and e_{it} is the emission by country i in period t. Payoffs are ongoing and period t payoffs are given by

$$h_i(e_{it}) - c_i g_t, \tag{24}$$

where $h_i(e_i)$ is a strictly concave utility function ("GDP") for country i. It represents the costs and benefits of producing and using energy (that then produces emissions e_i). The damage cost caused by GHG is linear and equals $c_i g_t$ where $c_i > 0$ is a constant marginal cost.

For a sequence of energy choices, we get the associated sequence of GHG stocks through Eq. 23. Then, lifetime payoffs for country i are given by

$$\sum_{t=0}^{\infty} \delta^t [h_i(e_{it}) - c_i g_t]. \tag{25}$$

A Simplification of Lifetime Payoffs - That the transition function, Eq. 23, is linear in the stock g and so is the stage payoff, Eq. 24, implies that lifetime payoff to energy usage e_i at date t can be computed on a stand-alone basis; separable from the GHG stock at t, emissions of other countries at t and emissions by country i in subsequent periods. The associated *lifetime payoff to country i* from energy usage e_i at date t is given by

$$h_i(e_i) - \delta w_i e_i, \tag{26}$$

where $w_i = \frac{c_i}{1-\delta\sigma}$ and the associated *lifetime cost to country $j \neq i$* from that same action, denoting $w_j = \frac{c_j}{1-\delta\sigma}$, is

$$-\delta w_j e_i. \tag{27}$$

Eqs. 26 and 27 are derived as follows. Energy usage e_i gives an immediate $t-th$ period (GDP) benefit $h_i(e_i)$; that is the first term. It also adds emission e_i to period $t+1$ GHG stock via the transition equation, Eq. 23, σe_i in period $t+2$, $\sigma^2 e_i$ in period $t + 3$, and so on. Given linearity in cost, Eq. 24, the marginal GHG cost is independent of other countries' energy choices and subsequent choices of country i; it equals $c_i \delta$ in period $t + 1$, $c_i \delta^2 \sigma$ in period $t + 2$, and so on. Hence, lifetime cost is $\delta \frac{c_i}{1-\delta\sigma} e_i \equiv \delta w_i e_i$. By identical arguments, e_i adds a per-period cost $\delta \frac{c_j}{1-\delta\sigma} e_i \equiv \delta w_j e_i$ to $j \neq i$.

The above argument implies that there is a stage-game, with no reference to the state variable g, and player $i's$ payoffs in the stage game are given by

$$h_i(e_i) - \delta w_i \sum_{j=1}^{N} e_j, \tag{28}$$

a function separable across other players' actions e_j and state g.

Suppose now that there is a cap $\bar{e}_i$ on total emissions that any country i can possibly generate. Abatement then can be written as $q_i \equiv \bar{e}_i - e_i$. Hence, Eq. 28 can be re-written as

$$g_i(q_i) - \delta w_i \sum_{j=1}^{N} \bar{e}_j + \delta w_i \sum_{j=1}^{N} q_j \qquad (29)$$

$$= \quad b_i \sum_{j=1}^{N} q_j - C_i(q_i),$$

where $b_i \equiv \delta w_i$, $C_i(q_i) \equiv g_i(q_i) - \delta w_i \sum_{j=1}^{N} \bar{e}_j$ and $g_i(\bar{e}_i - e_i) \equiv h_i(e_i)$. Of course, Eq. 29 is a heterogenous specification that can be turned into the homogeneous Barrett model by restricting $h_i(\bullet)$ and c_i to be country independent.

This static reduction of the Dutta-Radner model can then be used as a justification for the model studied in this paper. Of course, many other dynamic models of climate change will not have such a static reduction and in those cases, the static assumption made here will remain a constraint on the generality of the conclusions.

References

[1] Barrett, S., 2003. Environment and statecraft: The strategy of environmental treaty-making: The strategy of environmental treaty-making. OUP Oxford.

[2] Barrett, S., 2005. The theory of international environmental agreements. Handbook of environmental economics, 3, pp.1457-1516.

[3] Battaglini, M. and Harstad, B., 2016. Participation and duration of environmental agreements. Journal of Political Economy, 124(1), pp.160-204.

[4] Bernheim, D., and Whinston, B., 1986. "Common Agency," *Econometrica*, 54-4, 923-942.

[5] Buchanan, J. M., 1965. An economic theory of clubs. Economica, 32, 1–14.

[6] Carraro, C. and Siniscalco, D., 1993. Strategies for the international protection of the environment. Journal of public Economics, 52(3), pp.309-328.

[7] Chander, P., 2017. Subgame-perfect cooperative agreements in a dynamic game of climate change. Journal of Environmental Economics and Management, 84, pp.173-188.

[8] Chander, P. and Tulkens, H., 1995. A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. International tax and public finance, 2, pp.279-293.

[9] Chander, P. and Tulkens, H., 1997. The core of an economy with multi-lateral environmental externalities. International Journal of Game Theory, 26, pp.379-401.

[10] Coase, R., 1960. The Problem of Social Cost. Journal of Law and Economics, 3, 1–44.

[11] Dutta, P.K. and Radner, R., 2004. Self-enforcing climate-change treaties. Proceedings of the National Academy of Sciences, 101(14), pp.5174-5179.

[12] Dutta, P.K. and Radner, R., 2009. A strategic analysis of global warming: Theory and some numbers. Journal of Economic Behavior and Organization, 71(2), pp.187-209.

[13] Dutta, P. K. and Radner, R., 2023. The Paris Accord and the Green Climate Fund: A Coase Theorem. Working Paper.

[14] Dutta, P. K. and Siconolfi, P., 2024. Imperfect Monitoring in Asynchronous Games with Transfers. Working Paper.

[15] Harstad, B., 2016. The dynamics of climate agreements. Journal of the European Economic Association, 14(3), pp.719-752.

[16] Harstad, B., 2023. Pledge-and-review bargaining. Journal of Economic Theory, 207, p.105574.

[17] Heal, G., 1994. Formation of international environmental agreements. In Trade, Innovation, Environment (pp. 301-322). Dordrecht: Springer Netherlands.

[18] Hoel, M., 1992. International environment conventions: the case of uniform reductions of emissions. Environmental and Resource Economics, 2, pp.141-159.

[19] Nordhaus, W., 2015. Climate clubs: Overcoming free-riding in international climate policy. American Economic Review, 105(4), pp.1339-1370.

[20] Nordhaus, W., 2021. Dynamic climate clubs: On the effectiveness of incentives in global climate agreements. Proceedings of the National Academy of Sciences, 118(45), p.e2109988118.

[21] Prat, A., and A. Rustichini, 1998. Sequential Common Agency. WP, Tilburg University, Center for Economic Research.

[22] Prat, A., and A. Rustichini, 2003. Games Played through Agents. Econometrica, 71 (4), 989-1026.

Figure 1: Off path transfer (y axis) and sustainable club size (x-axis)