Is there a broader-deeper tradeoff?

Michael J. Gilligan

Abstract

Many scholars have argued that there is a tradeoff between the breadth and depth of multilateral institutions. I show that such a tradeoff does not exist for a broad class of multilateral cooperation problems. The conclusion that there is a broader-deeper tradeoff follows from the questionable, or at least not very general, assumption that the members of the multilateral must choose a single policy that is applicable to all members. The multilateral agreement modeled in this paper allows countries to set their policies at different levels. Once this change is made, there is no broader-deeper tradeoff, a finding which has obvious empirical and policy implications. It explains why some regimes are created with fairly large memberships at the outset, and it calls into question the policy prescription of limiting membership of multilateral institutions to a small group of committed cooperators for a large class of cooperation problems.

Acknowledgments

I am especially indebted to Marek Kaminski, Antonio Merlo and an anonymous reviewer for their help with the analysis in this paper. I would also like to thank Bruce Bueno de Mesquita, Adam Przeworski, Shanker Satyanath and the participants at seminars at UC-Berkeley and Rutgers University for comments on earlier drafts of this work. All errors remain my responsibility.
Is there a broader-deeper tradeoff?

**Introduction**

There has often been thought to be an inverse relationship between “broader” and “deeper” in international multilateral agreements—that is, multilateral agreements that are more inclusive will necessarily be shallower in their level of cooperation. In this paper I show that such a tradeoff does not exist for a broad class of multilateral cooperation problems. The conclusion that there is a broader-deeper tradeoff follows from a highly questionable (or at least not very general) assumption, namely that the members of the multilateral choose a single policy that is applicable to all members. The multilateral agreement modeled in this paper allows countries to set their policies at different levels. Once this change is made, there is no broader-deeper tradeoff. In other words, the broader-deeper tradeoff is a logical impossibility for the type of cooperation problem modeled in this paper. Indeed once the “single policy” assumption is relaxed, broader multilateral agreements can actually engage in deeper cooperation than narrower ones.

The findings in this paper have obvious empirical and policy implications. Empirically, models that generate a tradeoff between broader and deeper present a puzzle because many multilateral agreements are created with fairly large numbers of members, many of whom do not appear to be particularly committed to international cooperation (I will discuss some specific cases later in the paper). The approach I take here does a better job of explaining these cases. From a policy standpoint, the broader-deeper tradeoff is at the heart of policy recommendations in favor of adopting multilateral agreements with
relatively small exclusive memberships, at least in the early stages of the agreement. For instance Downs, Rocke and Barsoom (1998) claim that “large multilaterals that start out small will tend to become deeper in a cooperative sense than those that start out with many members (p. 398).” The analysis in this paper suggests that that statement is not true for a large class of cooperation problems. For the types of cooperation problems modeled in this paper excluding countries from the agreement only lessens the benefits of cooperation by reducing the number of countries that set their policies cooperatively while offering no increase (and perhaps even some reduction) in the depth of cooperation. Thus creators of multilateral agreements should in fact strive for inclusiveness in these cases.

A further purpose of this paper is to show that modeling multilateral bargaining is not so complex as to be avoided altogether. To date there has been little or no effort at modeling the creation of multilateral agreements formally. One reason for the lack of formal study of multilateral agreements is that the problem has been thought to be too complex analytically. One of the best formal studies of this topic avoids the multilateral bargaining issue stating: “The [multilateral institution] creation decision in its purest form is an unstructured bargaining problem of enormous complexity. For example in a world of fifty states there are over two million combinations by which a five state multilateral might form” (Downs, Rocke and Barsoom, 1998, p. 398). This paper will show that that statement makes the problem seem rather more complicated than it needs to be, and takes a first step at filling this gap in the formal international relations literature.

The paper is organized as follows. In the following section I will review a model of the creation of a multilateral organization in which members choose a single policy
level and show how that assumption leads to the conclusion that there is a broader-deeper tradeoff and that deeper may be better than broader. In section III, I will present my model of multilateral bargaining and the major findings that follow from it. Section IV will discuss some case examples that illustrate the findings. Section V concludes.

A model of voting on a single policy level

In this section I review a model of multilateral decision-making in which the members of the agreement vote on a single policy level for all members. A more formal treatment of this case, including proofs for the claims that I mention here, is available in Downs, Rocke and Barsoom (1998). Figure 1 illustrates the policy space. The left side of the dimension corresponds to more cooperative policy. We might think of the space as a continuum of the amount of pollution countries pump into a common body of water with the left side corresponding to lower levels of pollution. I have depicted a world of eleven countries. The claims obviously generalize to a world of more countries. Countries’ ideal points for the policy in question are arrayed along a continuum—$x_1$ is country 1’s ideal point, $x_2$ is country 2’s ideal point and so forth. Members of the organization vote on the level of policy that they want set for the group as a whole. With simple majority voting the equilibrium policy, in the sense that it cannot be overturned by a majority, will be the ideal point of the median of the admitted members. If all the countries are allowed to become members the cooperative policy would be country six’s ideal point. If instead a subset of countries, say countries one through five, created a more exclusive agreement the cooperative policy would be country three’s ideal point—much deeper cooperation than if the full group were admitted. This is the basis for the recommendation...
that multilateral agreements should be relatively exclusive in their early stages and admit members only as they become more willing to set their policies at a low level such as $x_3$.

Figure 1: Voting on a single policy with eleven countries

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
<td>$x_5$</td>
<td>$x_6$</td>
<td>$x_7$</td>
<td>$x_8$</td>
<td>$x_9$</td>
<td>$x_{10}$</td>
</tr>
</tbody>
</table>

Obviously of and by itself this argument does not definitively show that “deeper is better than broader.” A three country agreement among countries 1,2, and 3 offers deeper cooperation but completely forgoes cooperation among the remaining eight countries in the world. It is not necessarily true that the countries of the world—even country 2 whose ideal point is chosen in such a treaty—are better off with a treaty among these three members rather than all eleven. To make the claim that “deeper is better than broader,” one must add additional assumptions. Downs, Rocke and Barsoom (1998) for instance add the additional assumption that the countries’ ideal points are necessarily declining (in a predictable way) over time, so that countries with high ideal points now can be included when their ideal points fall. Regardless of the plausibility of this assumption, the argument must also address the question of why countries cannot adopt one agreement now and another agreement with deeper levels of cooperation later when countries’ ideal points have declined. This seems prima facie like a superior approach to dealing with changing ideal points. These issues are not the main point of this paper though. My intent is not to improve a model of voting on a single policy, but instead to alter these two assumptions (i.e. voting and the setting of a single policy).
The second of these assumptions is more important to the broader versus deeper debate, so that is where I will begin my discussion. Multilateral agreements do not generally require that their members set their policies at exactly the same level, as is assumed by the model I just discussed. For example membership in the Montreal Protocol did not require that each country set the same level of ozone depleting substances. The United States’ level of pollution was different from Japan’s, which was different from Canada’s level and so on. Similarly the Kyoto protocol sets different levels of greenhouse gas emissions for each of the members.

This single-policy assumption produces odd side effects. Presumably, prior to cooperation, countries were setting their policies at their national ideal points. One implication of this model is that countries that have ideal points below the median of the members of the multilateral will actually have to increase their policy level to be in compliance with the treaty. Even in the more exclusive treaty containing only countries one through five, countries one and two will actually have to increase their policy level up to \( x_3 \) in order to be compliant with the treaty. For such countries, being a member of the multilateral actually makes their policy less cooperative than not being a member.

The single policy assumption is crucial to the conclusion that there is a broader-deeper trade-off and that deeper is superior to broader. That conclusion is an artifact of the assumption that adding a less cooperative member increases every member’s policy level because it raises the median policy level. As I will show in the next section if the multilateral allows countries to set different policies, there is no broader-deeper tradeoff.
The second assumption that might be questionable is the assumption of voting. In the international context bargaining models are more appropriate than voting models. (In fact some models of policy making in domestic legislatures within countries treat it as a bargaining problem. Baron and Ferejohn (1989a) is the seminal piece in that literature.) Adopting a legislative voting approach when modeling policy making in international multilateral institutions assumes away two important problems in the creation of international multilateral agreements. The first of these concerns countries’ incentive to participate in the agreement and the second concerns enforcement. Legislative districts (e.g. states and Congressional districts in the US) that lose the vote on a policy issue must still abide by that policy. They cannot choose to opt-out of the policy. Countries that lose a “vote” in a multilateral organization, on the other hand, can simply refuse to sign the agreement, thereby dragging out the negotiations or, if the agreement enters into force anyway they can refuse to sign it. In that case they will not be bound by its provisions. Furthermore even if a country does sign the agreement it can cheat. Since there is no international authority to force states to participate in and comply with an international agreement that agreement must offer its members sufficient incentive to sign it and it must be self-enforcing. The model presented in the next section will explicitly address the problems of participation and enforcement.

There are other problems with adopting a legislative voting approach to international multilateral institutions which I do not address explicitly in this paper. One of the more important ones is that voting does not take into account differences in bargaining leverage. For example, if all of the countries in Figure 1 do not have the

---

1 Downs, Rocke and Barsoom (1998), for instance, assume that the cooperative policy level is set by a two-thirds majority vote.
power to propose treaties (this would be the case if proposals were sufficiently costly to devise and guide through the international negotiation process, for instance) then the policy adopted may not be the median position.\(^2\) As an empirical example, although developing countries in the GATT/WTO vastly outnumber the US and the EU, it is the US and the EU that generally determine most of the provisions of the treaties in that organization. I will not directly address this issue in this paper. In the model presented in the next section countries have equal bargaining power.

In summary, the assumption that countries in a multilateral agreement choose their cooperative policies via voting on a single policy level is not general. It may have its place in certain contexts, such as policy making within the European Union where countries really do often adopt a single policy for the whole Union (the establishment of a single currency being the most prominent example), and decisions are often made by qualified majority voting, but it is not a close match for many multilaterals. In the next section I present a model of multilateral bargaining in which countries are allowed to set different levels of the policy in question and the cooperative solution is arrived at via bargaining. In that model the broader-deeper tradeoff disappears.

The model

Some complications arise in the multilateral setting that are not present in bargaining over a bilateral international agreement. One of the more important complications is whether the benefits of cooperation are excludable or not. By excludable benefits I mean that non-signatories can be prevented from receiving the benefits of cooperation among

\(^2\) This is a common result. See, for instance, Baron and Ferejohn (1989b). Bueno de Mesquita and Stokman (1994) provide examples within the European Union context using a different model.
signatories. The key to excludability is whether or not a country can target its policies toward other countries. Some types of policies can be set differently toward different countries. Trade policy is one example. Country A’s tariff rate on products from country B can be different than its tariff rate on products from country C. Extradition treaties and treaties on the granting of visas are other examples. Policies that are targetable in this way will allow members of a multilateral treaty to exclude non-members from the benefits of cooperation because they can have different (presumably less cooperative) policies toward non-members than they do toward members.

In some cases, however, a country can only set its policy at one level. A country can only emit one amount of CFCs, for instance. Country A cannot emit one level of CFCs toward country B and a different level toward country C. Similarly, a country can only have one amount of biological weapons in its stockpile. In these cases members of a multilateral agreement will not be able to set different policies toward members than they do toward non-members. Non-members will be able to enjoy the benefits of cooperation even when they do not participate in the agreement.

For modeling purposes, excludable benefit treaties (i.e. cooperation over policies that can be targeted to specific countries) are quite different from non-excludable benefits treaties (cooperation over policies that cannot be targeted to specific countries). The two most important differences are 1) the way the agreements are enforced (i.e. the punishments for cheating on the agreement) and 2) whether countries that are invited to join the agreement have an unambiguous incentive to do so (or whether they have incentive to free ride). The first of these differences concerns whether or not punishments can be targeted to countries that cheat on the agreement. The latter of these differences
has important implications for what countries should do if some subset of them refuses to join the multilateral agreement and cooperate. In particular, it concerns the number of ratifications that are required for the treaty to come into force, in a way that I discuss below. These differences between the two types of cooperation problems are sufficient to require taking on each of them separately. In this paper I discuss a treaty in which benefits are non-excludable because I regard it as the harder case for international cooperation. The main result should go through for an excludable benefits treaty. We are now ready to turn to the model itself.

\( N \) is the set of countries, cardinality \( n \). Although it is not an essential feature of the model, substantively it makes sense to think of \( N \) as the set of countries that are relevant to the issue in question rather than all of the countries in the world. For instance an agreement to clean-up the Black Sea would have no reason to include all of the countries in the world. The agreement is only relevant to countries that border that body of water. Even in the case of an agreement that regulates a global issue, like protecting the ozone layer, \( N \) would not necessarily include all countries in the world because many countries’ consumption of ozone-depleting substances is negligible.\(^3\)

Each country \( i \in N \) sets a policy \( x_i \in \mathbb{R}^+ \). We can think of this policy as some “bad” like the amount of pollution produced by the country. Each country has an ideal point for its own policy, which I denote as \( z_i > 0 \). I assume that \( i \)’s ideal point for the other countries’ policies is zero. This assumption greatly reduces the amount of notation in the model and is not unreasonable substantively. If for instance we are discussing an environmental treaty it amounts to an assumption that countries do not want other

---

\(^3\) As I point out below the Montreal Protocol which regulated consumption of CFCs included only 29 countries at its signing but they made up 82% of global CFC consumption.
countries to pollute at all. Although there are cases that violate this assumption, it seems reasonable for many cases, does not change the substantive results and offers substantial gains in the tractability of the model. Each country’s utility is characterized by a quadratic loss function:

\[
U_i = -(x_i - z_i)^2 - \sum_{j \in N, i \neq j} x_j^2
\]  

(1)

Given this set-up, each country will set its own policy at its own ideal point \((z_i)\) and this outcome is globally sub optimal (this is a standard result, so I omit the proof). Each country would be willing to trade some reduction in its own policy for some reduction in other countries’ policies. To accomplish this the countries can agree to a treaty that reduces all of their policies. I will now discuss a game of bargaining over such a treaty, and the subsequent enforcement of the agreement.

GAME FORM: The game can be divided into two phases, a bargaining phase and an enforcement phase. Each player can take actions only at discrete times \(t \in \{1, 2, 3, \ldots\}\) which correspond to the stages of the relevant phase of the game. I will begin by describing the bargaining phase. “Nature” chooses a proposer from \(N\) according to some

---

4 One could imagine a case where country A exports a great to country B and therefore benefits when economic growth in country B is high. If reducing a certain pollutant reduces B’s economic growth, country A may prefer that that country B not reduce its level of this pollutant all the way to zero. An even starker example might be one in which country A’s and B’s pollution are caused by consumption of a product A imports from B, in which case a reduction in A’s pollution might have severe consequence for B’s exports. An empirical example of such a case might be Saudi Arabia and the Kyoto protocol (the global warming treaty).

5 Call countries \(i\)’s ideal point for \(j\)’s policy \(z_{ij}\). Then as long as \(z_{ij} < z_j\), the equilibrium policies would be higher (less cooperative) than those described in proposition 1 below, but still lower than in the non-cooperative equilibrium. The remaining conclusions would not change. If we allowed \(z_{ij} > z_j\), e.g. we assumed country \(i\) wanted country \(j\) to pollute more than country \(j\) wanted itself to pollute, it could affect the results more significantly.
rule (more on this rule later). The chosen proposer then makes a proposal $\Pi_1$ that includes a list of countries to be invited along with cooperative policies, $c_i \geq 0$, for each of them. The proposal does not specify policies for uninvited countries (i.e. if $\Pi_1$ is accepted those countries may set their policies wherever they like.) The set of countries that the proposer invites is allowed to be empty. The remaining countries in $N$ (the “non-proposers”) either accept or reject $\Pi_1$. If one or more of these rejects $\Pi_1$ the proposal fails and bargaining proceeds to the next stage (stage 2), and the stage game repeats. Nature picks the next proposer from $N$ according to the rule. The proposer that nature has just chosen then makes a proposal $\Pi_2$ that includes a list of the countries invited and their policies. This list can include countries that were not invited in the previous proposal. The non-proposers either accept or reject $\Pi_2$. If one or more of them rejects it, the game goes on to stage three. The bargaining stage game repeats in this way until all countries accept the proposal. Once this occurs, say at time $\tau$, the bargaining phase of the game ends and the proposal $\Pi_\tau$ is implemented. The game then proceeds to an enforcement phase in which the members of the agreement play a multilateral repeated prisoners’ dilemma. In each stage of this phase of the game members of the agreement either set their policies at those mandated by $\Pi_\tau$ or they do not (i.e. they cheat). If a member cheats the other members of the agreement punish the cheater by reverting to “punishment levels” of their respective policies for a specified period of time. (I will discuss these punishment levels in greater detail shortly.). After punishing the cheater for this specified

---

6 For substantive reasons we may want to limit the set of countries that can make proposals to some proper subset of $N$. Making this change complicates the game form described here and does not change the substantive results presented in this paper.
period of time countries return to their cooperative policies as specified by $\Pi_t$. Non-members are allowed to set their policies wherever they like without punishment.

To complete the game form, we must specify the rule by which a proposer is chosen. The empirical record does not suggest a clear answer to the question “Who proposes?” During the early post-war years the United States seemed to have a virtual monopoly on the power to propose. The most important post-war multilateral agreements—the GATT and the Bretton Woods system (Gardner, 1980), NATO (Osgood, 1962) and the UN (Hilderbrand 1990) were first proposed by the United States. A cursory review of the more recent history of the creation of multilateral agreements reveals that the US no longer plays such a dominant role. The question of who has first proposal power is an important one because as we shall see shortly first-proposal power confers certain advantages on the country that possesses it. Without more guidance from the empirical record I am going to simply assume that the proposer is chosen randomly and that each country has an equal probability of being chosen in any period.\footnote{We could modify the model to say that countries have a higher probability of being chosen the larger their power resources, for instance, but that is a relatively minor modification to the set-up I present here.} Note that the assumption implies that countries are allowed to make counterproposals to their own (previous) proposals.

I will now return to the issue of punishments. Since this is an international agreement the treaty must be self-enforcing. Each country must receive utility from complying with the treaty that is at least as high as it would receive from cheating on the agreement and being punished for it. The proposer’s problem is to propose a treaty that maximizes its utility subject to the constraint that all the signatories will comply with the
treaty. We must specify the punishments in order to be able to determine the compliance constraint. For simplicity, I assume the members of the treaty will punish cheating by returning to their Nash equilibrium policy levels (i.e. set their policies at their ideal points) for \( p \) periods. All of the countries have a common discount rate \( \delta \).

Note that in punishing a cheater the members of the agreement also “punish” each other. This is unavoidable given the non-excludability assumed here. Countries could in practice punish the cheater by setting their policies at something other than their non-cooperative Nash equilibrium policy—say some policy higher than the cooperative level but less than the full-blown non-cooperative level. Such a policy would only require a longer punishment period (\( p \)) in order to make the agreement enforceable, and of course if such a punishment is not sufficiently severe it is possible that the agreement will not be enforceable even for an infinitely long punishment period. I adopt the simpler assumption that countries return to their non-cooperative Nash equilibrium policies for the minimum number of periods that makes the agreement enforceable.

I would like to highlight one final feature of the game form. Agreement requires unanimity. If any country in \( N \) rejects the treaty no treaty is implemented in that period, and the bargaining continues. Without this feature some non-proposers may have an incentive to reject the treaty in order to free ride on the cooperation of those that signed the treaty. This is really a step-level public goods problem (Goeree and Holt 2003) with incentives similar to those of the well-known “paradox of (not) voting” (Ferejohn and Fiorina 1974). A sufficient number of countries must ratify the treaty for it to come into force.\(^8\) Countries may prefer that other countries ratify the treaty so that it would come

\(^8\) Ratification requirements vary from treaty to treaty. The treaty establishing the International Criminal Court required 60 ratifications before coming into force. Sometimes a “double trigger” is used. For
into force and they could free-ride on the cooperation of the other members, but if too
many of them play this strategy the treaty will not come into force in which case
countries will regret their decision not to ratify the treaty. The game form I have
described allows us to ignore these complications by effectively taking away the non-
proposers’ incentives to reject the treaty in order to free ride on the cooperation of the
other members. I should add that, in general, the proposer cannot solve this problem by
proposing a treaty that gives each non-proposer the utility they would receive from free-
riding because, in general, there are not sufficient gains from cooperation to do so and
still insure the compliance of the proposer.

Admittedly this assumption is a limitation of the model. In practice treaties do not
require that all relevant countries ratify them before coming into force, although they
usually do require that some substantial number of them do. Indeed the treaty’s
ratification requirement is typically part of the equilibrium. By that I mean it is decided as
part of the negotiations. Unfortunately, endogenizing the size of the ratification
requirement complicates the model considerably and is beyond the scope of this paper. I
would like to stress that this feature of the game form does not drive the main result of
the paper—namely that there is no broader deeper tradeoff. As we shall see shortly the
equilibrium cooperative policies in this model are a non-increasing function of the
number of member in the treaty, and this result does not depend on this assumption.
Finally I would like to stress that all of these problems are only present when the benefits
of the agreement are non-excludable. When benefits are excludable there is no incentive
to free ride and this assumption is unnecessary.

instance the Kyoto Protocol required ratifications of 55 countries comprising 55 percent of the 1990 global
CO₂ emissions. Similarly, the Montreal Protocol required 11 ratifications from countries comprising at least
two-thirds of 1986 CFC emissions.
To continue with the model let us call the set of countries in the treaty $M$, cardinality $m \leq n$. Compliance of $i$ requires:

$$-(1+\delta+\ldots+\delta^p) \left( (c_i-z_i)^2 + \sum_{j \in M} c_j^2 + X \right) \geq \sum_{j \in M} c_j^2 - X - (\delta+\ldots+\delta^p) \left( \sum_{j \in M} z_j^2 + X \right)$$

for all $i \in M$. Simplifying the compliance constraint yields:

$$-(c_i-z_i)^2 + \rho \sum_{j \neq i} (z_j^2 - c_j^2) \geq 0$$

(2)

where $\rho = \frac{\delta+\ldots+\delta^p}{1+\delta+\ldots+\delta^p}$.

As in other models of international cooperation, if the “shadow of the future” is not long enough there will not be sufficient gains from cooperation to insure the compliance of the members of the agreement. In this model the shadow of the future is captured by $\rho$ which is a combination of the discount rate and the number of punishment periods ($\rho$ acts like the usual discount rate, $\rho(p) < \rho(p+1)$ and $\lim_{p \to \infty} \rho = \delta$). I will assume that $\rho$ is not “too low” for the remainder of the paper.

The chosen proposer will want to insure not only that the other countries will comply with the treaty it proposes but also that they will accept it. To accomplish this goal the chosen proposer must offer the other countries their “continuation values”—that is, their expected utility of rejecting the proposal and taking the negotiations to the next stage. Define $V_{ij}$ as this continuation value for country $j$ of the treaty proposed by country $i$. $W_j$ is the value that country $j$ receives from an agreement that it proposes. We can
determine the continuation value as follows. If country $j$ is to accept the proposal of country $i \neq j$, then the following must be true.

\[
\frac{V_j}{1 - \delta} = Q_j + \frac{\delta}{n} \left( W_j + \frac{\sum_{k \neq i} V_k}{1 - \delta} \right)
\]  

(3)

where $Q_j$ is country $j$’s utility from the non-cooperative equilibrium. Equation (3) says if country $j$ is to accept country $i$’s proposal the net present value of that agreement taken indefinitely into the future (the left side of (3)) must at least be equal to the expected value that $j$ would receive from delaying the negotiations one more period (the right side).\(^9\) Country $i$ (the proposer) must offer $j$ at least this much in order to induce $j$ to accept the proposal. There will be an equation like (3) for each country indexed by $k$ in equation (3). These equations along with equation (3) form a system of equations that can be solved recursively to yield country $j$’s continuation value solely as a function of $n$, $\delta$, $Q_j$ and $W_j$. We will call this continuation value $V_j^*$. Completing this process shows that $V_j^*$ is clearly less than $W_j$ because it is a weighted average of $W_j$ and $Q_j$, which is less than $W_j$. Of particular importance is that $V_j^*$ does not depend on the identity of the proposer.

Country $j$’s per period utility from $i$’s proposed agreement must be at least equal to $V_j^*$ if $j$ is to accept the agreement. In other words:

\[
-(c_j - z_j)^2 - \sum_{\substack{k \in M \atop j \neq k}} c_k^2 - X - V_j^* \geq 0
\]  

(4)

\(^9\) Using $\delta$ for the discount factor in both the bargaining and enforcement phases, as I do here, implies that the stage game in each of these phases is the same length of time. I adopt this assumption to conserve on notation but it is easily altered and obviously has no effect on the results.
We will call equation (4) country $j$’s “continuation constraint.” Notice that since $V_i^* < W_i$ the proposer’s own continuation constraint will not bind. The proposer’s compliance constraint may bind of course. Notice also that the proposer will want to insure that all countries’ continuation constraints are met because if they are not countries will reject the proposal and the bargaining will continue to the next stage in which case the proposer’s expected utility is $V_i^* < W_i$.

If $\delta$ is low and there are many possible proposers ($n$ is large) $V_j^*$ may not be sufficient to keep country $j$ compliant, in which case the proposer will have to choose policies that satisfy country $j$’s compliance constraint as defined in equation (2). More formally the proposer must choose policies for country $j$ such that:

$$
(z_j - c_j)^2 \leq \min \left\{ \rho \sum_{\substack{k \in M \setminus j \neq k}} (z_k^2 - c_k^2), -V_j^* - \sum_{\substack{k \in M \setminus j \neq k}} c_k^2 - X \right\}
$$

Let’s call the group of countries for which the compliance constraint binds $A$ and the group of countries for which the continuation constraint binds $B$. The proposer will choose a proposal so as to maximize its utility subject to the compliance and continuation constraints. That is the chosen proposer will propose the agreement that maximizes:

$$
L_i = -(c_i - z_i)^2 - \sum_{j \neq i} c_i^2 - \sum_{j \in A \setminus i} \lambda_j \left( (c_j - z_j)^2 - \rho \sum_{\substack{k \in M \setminus j \neq k}} (z_k^2 - c_k^2) \right) - \sum_{j \in B} \lambda_j \left( (c_j - z_j)^2 + \sum_{\substack{k \in M \setminus j \neq k}} c_k^2 - X - V_j^* \right)
$$

where the $\lambda$s in (5) are Lagrange multipliers on the constraints in (2) and (4). We can now characterize the equilibrium of the game.
PROPOSITION 1: The equilibrium to the game described above has the following features.

a) Equation (5) has a unique maximum.

b) The proposer’s cooperative policy will be:

\[ c_i = \frac{(1 + \lambda_i)z_i}{1 + \lambda_i + \rho \sum_{j \in A} \lambda_j + \sum_{j \in B} \lambda_j} \]  

and the cooperative policy of the other countries \( j \in M \) will be:

\[ c_j = \frac{\lambda_j z_j}{1 + \lambda_j + \rho \sum_{k \in A} \lambda_k + \sum_{k \in B} \lambda_k} \]  

\[ (6) \]

\[ (7) \]

c) Bargaining will end in the first period

d) The agreement will include all \( N \) countries.

e) All members will comply with the agreement.

Proof: Please see the appendix.

It may be worthwhile to devote some space discussing two important aspects of this equilibrium. First with regard to the equilibrium policies, the complementary slackness conditions require:

\[ \lambda_j \left( (c_j - z_j)^2 - \rho \sum_{k \in M} (z_k^2 - c_k^2) \right) = 0 \]

for all \( j \in A \), and
\[ \lambda_j \left( (c_j - z_j)^2 - \sum_{k=1, k\neq j}^{M} c_k^2 - X - V_j^* \right) = 0. \]

for all \( j \in B \). These conditions say that the relevant constraint binds with equality or it does not bind at all, in which case the country’s Lagrange multiplier is zero. These conditions along with equations (6) and (7) imply three features of the equilibrium:

a) the proposer’s cooperative policy never reaches zero (i.e. the proposer never has to completely reduce its level of this “bad” to zero),

b) if the relevant constraint (compliance or continuation) is not binding on some country \( j \), the proposer puts forward a policy for \( j \) of zero (i.e. the proposer proposes that country \( j \) reduce its level of the “bad” to zero), and

c) if the relevant constraints of all members are non-binding, the proposer’s “cooperative” policy is its ideal point (i.e. if country \( i \) is the proposer, it proposes that its policy be \( z_i \) and all other countries policies be zero).

Examples of treaties like the one described in (c) do not spring easily to mind suggesting that, if this model matches reality, constraints typically do bind. The important point, though, is that features (a) through (c) highlight the benefits of being the proposer.

Second, part (d) of the equilibrium is interesting because according to the game form:

a) proposers are allowed to be “exclusive” in the sense that they are allowed to refuse to invite countries that are not committed to cooperation and

b) some countries may prefer to be excluded so that they can free ride on the cooperation of the other members rather than sign the treaty,
and yet part (d) of the equilibrium follows. While the game form requires that all countries accept the proposal it does not require that the proposer invite all countries, and some non-proposing countries would accept (indeed would prefer) a proposal in which they were not invited so that they could free ride on the cooperation of the members of the agreement. Countries in this latter group accept the proposal because doing so offers them the same utility, in expectation, as rejecting it, $V_i^*$. In short rejecting the treaty does not allow countries to free ride because it does not lead to a treaty that includes all other countries except them. Instead it only prolongs the negotiations.

It will be convenient in the pages that follow to refer to each country’s cooperative policy as a percentage of its non-cooperative equilibrium policy, $z_i$. Thus the proposer’s policy will be written:

$$c_i = \alpha_i z_i \text{ where } \alpha_i = \frac{1 + \lambda_i}{1 + \lambda_i + \rho \sum_{j \neq i} \lambda_j + \sum_{i \in B} \lambda_j}, \text{ and}$$

a generic non-proposing country’s policy will be written:

$$c_j = \alpha_j z_j \text{ where } \alpha_j = \frac{\lambda_j}{1 + \lambda_j + \rho \sum_{k \in A \atop j \neq k} \lambda_k + \sum_{k \in B \atop j \neq k} \lambda_k}.$$

We are now prepared to develop the main proposition of the paper, namely that there is no broader deeper tradeoff in the situation described by this model. Let us call the equilibrium policy of country $i$ from an agreement with $m$-1 other countries $\alpha_i(m)$, and without loss of generality let us call the first period proposer country 1 for the remainder of the paper. The next proposition shows that there is no broader-deeper tradeoff in this model.
PROPOSITION 2: $\alpha_i(m) \geq \alpha_i(m+1)$, $i \in \{2,3,\ldots,m\}$

Proof: Please see appendix

Proposition 2 simply states that the policies of the non-proposing countries proposed by the treaty will not be larger (less cooperative) for a treaty with more members. Adding an extra member can only change a member’s policy by lowering it. In this sense, broader treaties can actually be deeper. Proposition 2 says nothing about the effect of adding more members on the proposer’s policy. In some specific circumstances the proposer’s policy may actually be higher from a more inclusive treaty than a less inclusive one.\(^{10}\)

The intuition for the absence of a broader-deeper tradeoff is quite straightforward, and is perhaps best explained with an example. Imagine a case of three countries—call them A, B and C—that are deciding to create a multilateral agreement to reduce pollution of a common body of water. Two of the countries A and B are quite committed, but country C is not at all committed to this environmental project. If countries A and B can come to an agreement with relatively low amounts of pollution between them why should adding the third country in any way require them to increase their pollution from that level? Increasing their (A’s and B’s) pollution in no way makes the agreement more palatable to country C—on the contrary. The third country wants to keep its own

\(^{10}\) For example, consider the case of an \(m\)-member treaty in which all countries’ policies are zero except that of the proposer and country 2. In other words only country 2’s constraint (whether it is the compliance constraint or the continuation constraint is immaterial) binds. We now consider adding country \(m+1\) to the treaty, and country \(m+1\) also has a cooperative policy of zero (neither constraint is binding on it), and suppose that by adding country \(m+1\) country 2’s constraint no longer binds. In that case country 2’s cooperative policy would fall to zero, the policies of countries 3 through \(m\) would remain zero, but the proposers cooperative policy would actually increase to \(z_i\).
pollution levels high; it does not want the other two countries’ pollution levels to be high. An agreement in which the two “environmentalist” countries set relatively low pollution limits and the third country sets a relatively high pollution limit (but still lower than the pre-treaty level) makes all three countries better off.

The broader-deeper tradeoff is an artifact of the often unspoken assumption that all members of a treaty must set their policy at the same level, so that a lower level of cooperation (e.g. more pollution) for everyone is required to coax the more intransigent countries into the agreement. But, this assumption is not applicable to many multilateral institutions and as such should not guide our prescriptions about the proper size of multilateral institutions.

Examples

A brief review of the creation of a few multilateral agreements reveals that, in many cases, issues of broader versus deeper simply did not arise. Certainly the proposer of some of our most important multilaterals in the early postwar era (the United States) was concerned with making them as expansive as possible, not with limiting their membership to “hard core” members. Kahler (1992, 681) put it this way: “Postwar multilateralism … expressed an impulse toward universality that implied relatively low barriers to participation in these instruments.”

The creation of the Bretton Woods system is an example of an important post-war institution in which you find no references to a “broader versus deeper” debate. Indeed the framers of the system hoped that it would be as expansive as possible. The framers did not seem to be concerned with restricting membership to those countries that
supported a stable fixed exchange rate system (Block, 1977, Gardner, 1980). The IMF opened its doors in 1946 with 39 members, about 89 per cent of the world’s market economies at the time, and an even larger percentage global GDP or world trade.

The GATT was created with 27 members, not a large organization by today’s standards, but it comprised over 60 percent of the countries with market based economies in 1947, and a much larger share of world trade. The U.S. invited countries widely. No country was excluded by the treaty’s core members because they thought those countries were too protectionist and would limit the depth of cooperation of the other members. Indeed some early members of the GATT were some of the most protectionist countries in the world (India, which was admitted upon its independence in 1948, is a good example). Much of the “evolution” of the GATT—meaning the addition of countries after the forming of the agreement—involves adding former colonies as soon as they became independent. Most of the growth of the membership of the GATT during the 1950s and 60s was pro forma as former colonies of GATT members were given easy membership in the organization.11

A more recent example might be the Montreal Protocol, which limited emission of ozone-depleting CFCs. The treaty came into force in 1989 with a membership of 29 countries and the EC, representing 82 percent of world CFC consumption (Benedick 1991). As of December 2002 the agreement had 184 members. A similar example is the Kyoto protocol which is meant to reduce CO2 emissions. That agreement received its 100th ratification in December 2002, and awaits the ratification of the Russian Federation.

11 We do have to be cautious in interpreting the history of trade policy cooperation in terms of the model presented here, because trade policy cooperation is clearly a case where benefits are excludable, but I expect the results to translate to excludable benefits treaties although I have not actually modeled it.
One final example is the Anti-Personnel Land Mines ban, which was opened for signature in December 1997 and to date has received 133 signatures. It was the stated intention of Lloyd Axworthy the Canadian Minister of Foreign Affairs who was the driving force behind the treaty to make the agreement “open to all, hostage to none (Axworthy and Taylor, 1998, p. 196).” Of course the US did not sign the land mines ban but this fact is deceptive since the US made it clear it would abide by the principles of the ban except the Korean peninsula and it was the largest contributor to land mine clean-up.

We can compare these institutions to one that seems to meet the “single policy” assumption more closely—the European Union. The European Union often sets a single policy that will apply to all countries in the Union and will be implemented more or less uniformly by a centralized agency. In these cases countries are not allowed to set different levels of the policy, and in some cases they relinquish control of the policy to a supranational authority. For instance, the Union (then Community) has set external trade policy as a unit since the late 1950s, which required that some countries actually raise their tariffs on trade with countries outside the EC. Indeed this was the cause for some trade frictions between the U.S. and the European Community as some members of the EC raised tariffs on some products (notably frozen chicken) to comply with the Common External Tariff (CET). The U.S. argued that as countries raised their tariff rates on frozen chicken in order to comply with the CET, they were violating their GATT commitments. The U.S. retaliated, starting “The Chicken War.” (Conybeare, 1987)

---

12 As mentioned above the Kyoto protocol employs a “double trigger” to come into force. Ratification by 55 countries comprising 55 percent of 1990 greenhouse gas emissions is required for the treaty to come into force. Approval by the Russian Federation with its 17.4 percent of 1990 greenhouse gas emissions will bring the treaty into force. As of this writing the Russian Federation is expected to ratify the treaty in early 2003.
We can also tell a story about the creation of the European Central Bank and European monetary union with the “single policy” framework. This may be a case where such an assumption is appropriate because a single money supply is set for all countries in the EU. Issues of “broader versus deeper” arose in the creation of the monetary union. There were discussions of a “multi-speed” Europe in which a “hard core” of countries would proceed with monetary union while the “soft core” countries waited. Furthermore the stringent conditions placed on countries’ macroeconomic policies (size of budget deficit, inflation and so on) before they could join the union might be interpreted as an attempt to exclude countries that would have had “wrong” preferences on macroeconomic policy. In short, the EU provides an example where it looks like the single policy assumption is appropriate, there were concerns about “broader versus deeper” and some potential members were effectively excluded as a result of it.\(^{13}\) The point of this paper is that these conditions are not general and indeed it was fairly easy to come up with examples where these conditions were not met. The point of this paper is that the single policy assumption and the tension between broader and deeper that it generates is not appropriate for the general class of multilateral agreements. As such the normative conclusion that those treaties should be exclusive does not follow.

\(^{13}\) I should point out that this is not the last word on policymaking within the EU. Recent work has questioned the broader versus deeper tradeoff even in the EU context, and of course cooperation has actually deepened in the EU even as its membership has broadened. Pahre (1995) offers a model in which a broader Union produces deeper cooperation. His is a model of joint provision of public good (really a club good; benefits are excludable). Enlarging the membership increases the number of countries that can contribute to the cooperative provision of this good, thus “deepening” cooperation in the sense of greater provision of this good.
Conclusion

The result that a tradeoff between broader and deeper exists flows from the assumption that the members of the multilateral must set their policy at identical levels. For many very important multilateral agreements this assumption is not appropriate, because countries are allowed to set their policies at different levels. In multilateral agreements in which each country sets a different policy level the broader-deeper tradeoff does not exist, and the policy implication that membership in multilateral agreements should be exclusive to a “hard core” of committed cooperators does not follow.

The model in this paper explains the historical patterns of the creation of some of our most important multilaterals better than the “narrower and deeper is better” model. In particular it explains why the U.S. was interested in making the GATT and the Bretton Woods system as expansive as possible. It explains the same pattern in negotiations over more recent multilateral agreements ranging from the Montreal and Kyoto Protocols to the Anti-Personnel Land Mines Ban. It also explains why those agreements included (or at least attempted to include) countries that did not appear to be particularly committed to the principles of liberal world trade, stable non-interventionist exchange rate policy, environmentalism and so on.

 Obviously, there is still much work to be done on bargaining over multilateral agreements. I make no claims about the generality of the game form presented in the paper. Indeed it would be impossible to do so because so little is known about the “general” case of bargaining over multilateral agreements. For this reason, empirical work on how the agreements were negotiated will be very useful. Beyond the few most famous agreements, we do not have a great deal of data on which countries proposed the
multilateral agreements, who made counterproposals, how long bargaining over them took, and so on. Gathering such data will be time consuming, but very important.

My hope in submitting this paper at this stage is to accomplish three goals. First I would like to convince the discipline that modeling multilateral bargaining is not so difficult as to be avoided altogether and to offer a simple model as a starting point. Second I would like to call into question the appropriateness of the single policy assumption for many important multilaterals. Finally I would like to caution that the policy recommendation that follows from that assumption, namely excluding countries that would generate substantial cooperative gains, is questionable.

**Appendix: Proofs of Propositions**

**PROPOSITION 1:** The equilibrium to the game described above has the following features.

a) Equation (5) has a unique maximum.

b) The proposer’s cooperative policy will be:

\[
c_i = \frac{(1 + \lambda_i)z_i}{1 + \lambda_i + \rho \sum_{j \in A, i \neq j} \lambda_j + \sum_{j \in B} \lambda_j}
\]

(6)

and the cooperative policy of the other countries \( j \in M \) will be:

\[
c_j = \frac{\lambda_j z_j}{1 + \lambda_j + \rho \sum_{k \in A, j \neq k} \lambda_k + \sum_{k \in B, k \neq j} \lambda_k}
\]

(7)

c) Bargaining will end in the first period

d) The agreement will include all \( N \) countries.

e) All members will comply with the agreement.
PROOF: I will prove each part of the proposition in turn.

a) Define $E$ as the set of all points $(x_1, x_2, \ldots, x_m)$ such that $0 \leq x_i \leq z_i \ \forall i \in M$. $E$ is a compact subset of $\mathbb{R}^m$. The constraints define a closed subset of $E$, $E^*$, such that all the constraints are satisfied and $E^*$ is nonempty (we know the constraints are satisfied at the Nash equilibrium, for example). Thus a maximum exists. Furthermore since $L_i$ is monotonically increasing in $c_i$ and monotonically decreasing in $c_j, j \in M, i \neq j$, that maximum is unique.

b) Equations (6) and (7) are derived from the first order conditions for a maximum to (5).

c) By construction all members of $B$ will accept the first period proposer’s offer because it gives them the same expected utility they would receive from rejecting that proposal. For countries in $A$ the continuation constraint does not bind but the compliance constraint does. This could only be the case if

\[\rho \sum_{j \neq i} (z_j^2 - c_j^2) < -V_i^* - \sum_{j \neq i} c_j^2 - X.\]

Thus

\[(1 - \alpha_i)z_i^2 = \rho \sum_{j \neq i} (z_j^2 - c_j^2) < -V_i^* - \sum_{j \neq i} c_j^2 - X,\]

and the proposed policies are more than sufficient to induce these countries to sign the treaty.

d) We have already shown that all invited countries will accept their invitations. We still need to show that the proposer will invite all countries. This is obviously true when $\lambda_j = 0 \ \forall j \in M$. For the remaining cases consider an $m+1$ member treaty in which the policies of countries 1 through $m$ are set at the equilibrium levels for an $m$-member treaty and the policy of country $m+1$ is set at the lowest level for which country $m+1$ will comply with
the treaty. The policies of the \(m\) countries in an \(m\)-member agreement are \(\alpha_i(m) \ i = \{1, 2, \ldots, m\}\). These policies insure the compliance of all \(m\) members and maximize the proposer’s utility. There exists an \((m+1)\) member treaty with which all members will comply and in which the policies of members 1 through \(m\) are \(\alpha_i(m) \ i = \{1, 2, \ldots, m\}\), and the policy of country \(m+1\) is \(\alpha_{m+1} x_{m+1}\) with an \(\alpha_{m+1} < 1\).

To see this first consider the case where country \(m+1\) is in A (i.e. the compliance constraint binds). Compliance of country \(m+1\) with such a treaty requires:

\[
(1 - \alpha_{m+1})^2 z_{m+1}^2 \leq \rho \sum_{i=1}^{m} (1 - \alpha_i^2(m))z_i^2 \quad (A1)
\]

Since the value on the right side of (A1) is strictly greater than zero there exists some \(\alpha_{m+1} < 1\) for which (A2) is true. Next consider the case where country \(m+1\) is in B. Participation of country \(m+1\) in such a treaty requires:

\[
(1 - \alpha_{m+1})^2 z_{m+1}^2 \leq -\sum_{i=1}^{m} \alpha_i^2(m))z_i^2 - X - V_{m+1} \quad (A2)
\]

Notice that the first two terms on the right side of (A2) equal the utility that country \(m+1\) would receive from free-riding on the equilibrium agreement among countries 1 through \(m\). Thus, if the utility from free riding on an agreement between countries 1 through \(m\) is greater than the utility from being in the equilibrium agreement with countries 1 through \(m\) the value on the right side of (A2) will be positive so there will be some \(\alpha_{m+1} < 1\) for which (A2) is true. I will discuss the case where this is not true below later in the proof.

Let’s call the value of \(\alpha_{m+1}\) that satisfies the relevant constraint (A1 or A2) \(\bar{\alpha}_{m+1}\). Note that I have made no claims about the optimality \(\bar{\alpha}_{m+1}\), only that country \(m+1\) would comply with/participate in such a treaty. Notice also that \(\bar{\alpha}_{m+1}\) may be zero. So far we have shown that there exists an \((m+1)\) member treaty with which a country \((m+1)\) that
meets the conditions described above will comply and in which the policies of members 1 through \( m \) are \( \alpha_i(m)z_i, \ i = \{1, 2, \ldots, m\} \), and the policy of country \( m+1 \) is \( \alpha_{m+1}z_{m+1} \) with an \( \alpha_{m+1} < 1 \).

Clearly, countries 1 through \( m \) will also comply/participate in with such a treaty. To see this consider country 2’s compliance constraint from an \( m \)-member agreement if country 2 is in A:

\[
(1 - \alpha_2(m))^2 z_2^2 = \rho(1 - \alpha_1^2(m))z_1^2 + \rho \sum_{i=3}^m (1 - \alpha_i^2(m))z_i^2
\]  
\[\text{(A3)}\]

With the addition of country \( m+1 \) setting its policy at the policy at \( \alpha_{m+1} \), country 2’s compliance constraint becomes:

\[
(1 - \alpha_2(m))^2 z_2^2 < \rho(1 - \alpha_1^2(m))z_1^2 + \rho \sum_{i=3}^m (1 - \alpha_i^2(m))z_i^2 + \rho(1 - \alpha_{m+1}^2)z_{m+1}^2
\]  
\[\text{(A4)}\]

If instead country 2 is in B then its continuation constraint from an \( m \)-member agreement is:

\[
(1 - \alpha_2(m))^2 z_2^2 = -V_2 - \alpha_1^2(m)z_1^2 - \sum_{i=3}^m \alpha_i^2(m)z_i^2 - z_{m+1}^2 - \sum_{i=m+2}^n z_i^2
\]  
\[\text{(A5)}\]

I have simply split \( X \) into two of its component parts in (A5). With the addition of country \( m+1 \) setting its policy at the policy at \( \alpha_{m+1} \), country 2’s continuation constraint becomes:

\[
(1 - \alpha_2(m))^2 z_2^2 < -V_2 - \alpha_1^2(m)z_1^2 - \sum_{i=3}^m \alpha_i^2(m)z_i^2 - \alpha_{m+1}^2 z_{m+1}^2 - \sum_{i=m+2}^n z_i^2
\]

Obviously, a similar argument applies to the remaining countries in the treaty.

To summarize, I have shown that there is an \((m+1)\) member treaty in which: a) the policies of countries 1 through \( m \) are kept at their levels from an \( m \)-member treaty and the
policy of country $m+1$ is set at a cooperative level less than its non-cooperative Nash
equilibrium level, and b) all countries will comply with this treaty, as long as the
conditions on country $m+1$ apply. Such a treaty increases proposer’s utility over that of
the $m$-member treaty by the amount $(1 - \tilde{a}_{m+1}^2)z_{m+1}^2$. If the treaty described here
unambiguously increases the proposer’s utility, then the treaty that maximizes the
proposer’s utility must also do so. Thus the proposer’s utility unambiguously increases
from adding the extra member.

We have yet to show that adding a country for which the free riding value is less
than the value of being in the treaty unambiguously improves the proposer’s utility. If the
proposer does not include such a country that country’s utility will be $V_i^* < W_i$. Therefore the proposer will include countries in this category as well.

e) This is obvious. The policies of countries in $A$ are chosen to insure their compliance.
For countries in $B$ their continuation constraints are more binding than their compliance
constraint so the former constraints are met, insuring the compliance of these countries. □

PROPOSITION 2: $\alpha_i(m) \geq \alpha_i(m+1)$, $i \in \{2, 3, \ldots m\}$.

PROOF: I will prove the proposition by contradiction. We only need to discuss the case
when $\alpha_i(m+1) > 0$, since if this condition does not hold the proposition is proved because
$\alpha_i(m) \geq 0$. I will proceed in three steps. The first of these will prove a result that is
necessary for the other two steps. The remaining two steps will complete the proof by considering two different cases for country \( m+1 \): 1) the case of a country in group B for whom \( \sum_{i=1}^{m} \alpha_i^2 z_i^2 \leq V_{m+1}^* \) and 2) all other cases, namely a country in group B for whom \( \sum_{i=1}^{m} \alpha_i^2 z_i^2 > V_{m+1}^* \) or a country in group A.

To prove the proposition suppose to the contrary that \( \exists \alpha_i(m) < \alpha_i(m+1) \), for some \( i \in \{2, 3, \ldots m\} \).

**Step 1:** If \( \alpha_i(m) < \alpha_i(m+1) \), for some \( i \in \{2, 3, \ldots m\} \), then \( \alpha_i(m) < \alpha_i(m+1) \), for all \( i \in \{2, 3, \ldots m\} \).

This statement is trivially true for all \( \alpha_i(m) = 0 \), because we are only discussing the case where \( \alpha_i(m+1) > 0 \). It is also true for \( \alpha_i(m) > 0 \), because the complementary slackness condition for countries in A requires:

\[
(1 - \alpha_i)^2 z_i^2 = \rho \left( \sum_{h=1}^{k} (1 - \alpha_h^2) z_h^2 \right) = (1 - \alpha_j)^2 z_j^2 - \rho \left( \sum_{h \neq j}^{k} (1 - \alpha_h^2) z_h^2 \right) = 0
\]

for any two countries \( i \) and \( j \) with non-zero cooperative policies. Rearranging we find that:

\[
\frac{(1 - \alpha_i)^2 + \rho(1 - \alpha_i^2)}{(1 - \alpha_j)^2 + \rho(1 - \alpha_j^2)} = \frac{z_j^2}{z_i^2}
\]

(A6)

The right side of (A6) is a constant, which means that any change in \( \alpha_i \) must be matched by a proportional change in \( \alpha_j \) in the same direction. A similar calculation using the
complimentary slackness conditions for countries in B shows that their policies must also move in the same direction.

**Step 2:** Country $m+1$ is in group B and $\sum_{i=1}^{m} \alpha_i z_i^2 \leq V_{m+1}^*$

This is the case of adding a country for which the utility of being in the treaty is greater than the utility of free-riding on an $m$-member treaty. Formally:

$$V_{m+1}^* = -(1 - \alpha_{m+1} z_{m+1})^2 - \sum_{i=1}^{m} \alpha_i z_i^2 - \sum_{i=m+2}^{n} z_i^2 > -\sum_{i=1}^{m} \alpha_i z_i^2 - \sum_{i=m+2}^{n} z_i^2$$

Simplifying this inequality shows:

$$(1 - \alpha_{m+1} z_{m+1})^2 \leq \sum_{i=1}^{m} \alpha_i z_i^2 - \sum_{i=m+2}^{n} z_i^2$$

Since the left side of this inequality must be greater than or equal to zero, the right side of the inequality shows that $\sum_{i=1}^{m} \alpha_i z_i^2 \geq \sum_{i=1}^{m} \alpha_i z_i^2$. This can only be the case if $\alpha_i z_i^2 = \alpha_i z_i^2 \forall i$, or if at least one of these $\alpha_i$ must be larger as a result of adding country $m+1$. But if one $\alpha_i$ is larger they all must be according to the argument presented above. This completes the proof for this case.

**Step 3:** Country $m+1$ is in group B and $\sum_{i=1}^{m} \alpha_i z_i^2 > V_{m+1}^*$ or country $m+1$ is in group A.

According to the proof of Proposition 1d, $U_1(m) < U_1(m+1)$ in all of these cases, a fact which we will use below.
We must consider two different sub-cases depending on whether or not country 1’s compliance constraint binds in the \( m+1 \) member treaty. In case (a) country 1’s compliance constraint binds for an \( m+1 \) member treaty, and in case (b) it does not bind for an \( m+1 \) member treaty.

Case (a)

The following condition holds for the \( m \)-member treaty:

\[
(1 - \alpha_i(m))^2 z_i^2 - \rho \sum_{i=2}^{m} (1 - \alpha_i^2(m)) z_i^2 \leq 0
\]

and for the \( m+1 \) member treaty:

\[
(1 - \alpha_i(m+1))^2 z_i^2 - \rho \sum_{i=2}^{m+1} (1 - \alpha_i^2(m+1)) z_i^2 = 0
\]

Subtracting the second of these conditions from the first and rearranging:

\[
\rho \left( (1 - \alpha_i(m))^2 - (1 - \alpha_i(m+1))^2 \right) z_i^2 + \sum_{i=2}^{m} (\alpha_i^2(m) - \alpha_i^2(m+1)) z_i^2 - (1 - \alpha_{m+1}^2) z_{m+1}^2 \leq
\]

\[
(1 - \rho) \left( (1 - \alpha_i(m+1))^2 - (1 - \alpha_i(m))^2 \right) z_i^2
\]

(A7)

The left side of (A7) is \( \rho \left( U_i(m+1) - U_i(m) + z_{m+1}^2 \right) \), which is strictly positive, by the proof of proposition 1d, and the right side of (A8) is strictly negative if \( \alpha_i(m+1) > \alpha_i(m) \). This contradiction completes the proof for case (a).

Case (b)

We know from the first order conditions that:
\[ \lambda_i(z_i - x_i) = \left( 1 + \rho \lambda_i + \rho \sum_{j \in A, i \neq j} \lambda_j + \sum_{j \in B, i \neq j} \lambda_j \right) x_i \]

We can combine the \( \lambda \)s on the right side and solve for \( \lambda_i \) to obtain

\[
\frac{\lambda_i}{1 + \rho \lambda_i + \rho \Lambda_A + \Lambda_B} = A_i \quad \text{(A8)}
\]

where \( \Lambda_A = \sum_{i \in A} \lambda_i \), \( \Lambda_B = \sum_{i \in B} \lambda_i \) and \( A_i \) is either \( \frac{\alpha_i}{1 - \alpha_i} \) if \( i \in A \) or \( \frac{\alpha_i}{1 - \alpha_i} \) if \( i \in B \).

Note that \( A_i \) is strictly increasing in \( \alpha_i \). We can sum the left and right side of (A8) over all \( i \in \{2, \ldots, n\} \) to obtain the following relations:

\[
\frac{\Lambda_A}{1 + \rho \lambda_i + \rho \Lambda_A + \Lambda_B} = \sum_{i \in A} A_i = A_A
\]

\[
\frac{\Lambda_B}{1 + \rho \lambda_i + \rho \Lambda_A + \Lambda_B} = \sum_{i \in B} A_i = A_B
\]

\[
\frac{\Lambda_A + \Lambda_B}{1 + \rho \lambda_i + \rho \Lambda_A + \Lambda_B} = \sum_{i \in M} A_i = A_{AB}
\]

By our earlier discussion if one of the \( A_i \)s on the right side increases with the addition of the \( m+1 \)th member all the \( A_i \)s do. Since, by assumption, \( \alpha_i(m) < \alpha_i(m+1) \), it must be true that \( A_j(m) < A_j(m+1), j \in \{A, B, AB\} \). This fact in turn implies \( \Lambda_A(m) < \Lambda_A(m+1) \) and \( \Lambda_B(m) < \Lambda_B(m+1) \). To see this note that \( \Lambda_A \) and \( \Lambda_B \) cannot both decrease with the addition of country \( m+1 \), because if they did it would imply a decrease in \( A_{AB} \). Furthermore \( \Lambda_A \) cannot decrease as a result of adding country \( m+1 \) because if it did while \( \Lambda_B \) increased (or did not change) it would imply a decrease in \( A_B \). A similar argument shows that \( \Lambda_B \)
cannot decrease while \( \Lambda_A \) increases (or does not change). Thus we know that \( \Lambda_A(m) < \Lambda_A(m+1) \) and \( \Lambda_B(m) < \Lambda_B(m+1) \).

Recall from the first order conditions that:

\[
\alpha_i(m) = \frac{1 + \lambda_i(m)}{\rho \lambda_i + \rho \Lambda_A(m) + \Lambda_B(m)} \quad (A9)
\]

and

\[
\alpha_i(m+1) = \frac{1}{\rho \Lambda_A(m+1) + \Lambda_B(m+1)} \quad (A10)
\]

\( \lambda_1 \geq 0 \) in (A9). Using equations (A9) and (A10), the fact that \( \Lambda_A(m) < \Lambda_A(m+1) \) and \( \Lambda_B(m) < \Lambda_B(m+1) \) implies that \( \alpha_i(m) > \alpha_i(m+1) \).

Our discussion so far has indicated that if it is true that \( \alpha_i(m) < \alpha_i(m+1) \) then the effect of adding country \( m+1 \) is to increase the policies of countries 2 through \( m \) and decrease the policy of country 1. By assumption this is the agreement that country 1 proposes in equilibrium (i.e. the agreement that maximizes country 1’s utility). Let us call this agreement \( \Theta(m+1) \) and the equilibrium \( m \)-member agreement \( \Theta(m) \). Consider instead an agreement that includes countries 1 through \( m+1 \) but in which the policies of country 2 through \( m \) are kept at the levels for an \( m \)-member treaty. Let us call this agreement \( \Omega \). If countries 1 and \( m+1 \) will comply with \( \Theta(m+1) \) they will surely comply with \( \Omega \) because \( \alpha_i(m) < \alpha_i(m+1), \ i \in \{2, \ldots, m\} \). If countries 2 through \( m \) will comply with \( \Theta(m) \) they will surely comply with \( \Omega \) because \( \alpha_1(m) > \alpha_1(m+1) \), and \( \alpha_{m+1}(m+1) < 1 \). \( \Omega \) offers country 1 unambiguously greater utility than agreement \( \Theta(m+1) \) does, but \( \Theta(m+1) \) supposedly maximized the proposer’s utility. This contradiction completes the proof for case (b) and proves Proposition 2. \( \square \)
Bibliography


