Sovereignty and Delegation in International Organizations

David Epstein and Sharyn O'Halloran[*]

1 Introduction

Scholars of international organizations have made great strides in recent years defining the crucial issues to understanding delegation relationships within these regimes. As Bradley and Kelley (2007) note elsewhere in this volume, the language describing international organizations now includes the notions of sovereignty costs, partial adoptions of international conventions and a variety of institutions exercising delegated authority. The vigorous discussions in this literature now revolve around the appropriate characterization of these terms and a clearer understanding of why some countries join organizations completely, partially, or not at all.

While this work proceeds apace, formal theoretic treatments of these issues have fallen behind. We argue that this is due to the fact that previous models of international organizations are fundamentally ill-suited to address these questions. One modeling tradition begun by Putnam (198X) sees international organizations as solutions to coordination problems that countries cannot solve efficiently through a series of bilateral agreements. A second, due originally to Tsebelis (199X) examines

the impact of domestic bargaining on the "two-level" games of international relations. And recently, a third line of inquiry has emphasized the informational advantages conferred by membership in international organizations.[1]

Each of these literatures highlights important aspects of cooperative organizations, but none can easily address issues of partial association, the size of international organizations, or the distribution of power within those organizations. Nor would the definitions of sovereignty costs and associational benefits be easily formulated within any of these frameworks.

The present essay seeks to remedy this gap by introducing a formal model of delegation to international organizations based on the complementary notions of free exit and network externalities. To begin with, what sets international organizations apart from countries' internal delegation regimes is the fact that if a country is not satisfied with the results they are obtaining via membership in the organization, they can simply exit. This stands in contrast to, for instance, interbranch delegation; if Congress is unhappy with the executive's use of delegated authority, they cannot simply declare themselves to be the new national legislature of Bolivia!

The glue that holds international organizations together, then, is not a mandate to legislate together, but rather network externalities. That is, a defining feature of such organizations is that the more countries that belong, the more benefits accrue to all members. So they display increasing returns to scale, similar to many social or internet-based resources. (cites)

Our new formal framework for the analysis of international organizations allows us, as described below, to talk naturally about delegation and limits on discretionary authority; about having countries join parts of an agreement without ratifying

---

[1]See Milner and Rosendorff (1996), Kydd (2003) and Johns (2006).

the agreement in its entirety; about sovereignty costs and associational banefits; about domestic political costs to joining international bodies; about allowing certain countries a veto over the organization's decisions; about the long-term viability of international organizations and their breakdown; and about the distribution of power within organizations. The purpose of the present paper is to introduce our model and elucidate some of its basic workings, while some of the more ambitious extensions are left for future work.

The remainder of the paper is organized as follows. The next section provides an overview of our modelling approach and relates it to the terms of debate in the sovereignty literature. The follow section provides a formal specification of our model. The forth section provides variance on this basic scheme and the last section concludes with implications for the long-run sustainability and optimal structure of international organizations.

2 Model Preview

The framework for analyzing delegation in political science is a variant on the "spatial model" of political institutions. This model assumes that actors have most-preferred policies, or "ideal points," in some policy space and want to bring policy outcomes as close as possible to their ideal point. It is political institutions that determine who gets enfranchised into the policy making process, under what conditions, and thus who wins and loses at the end of the day.

This setup is illustrated in Figure 1. The small (turquoise) dots represent the ideal points of member states in some international organization. The large (blue) dot indicates the status quo policy *ex ante*. The policy making institutions of the organization determine in which way policy will move in response to various external
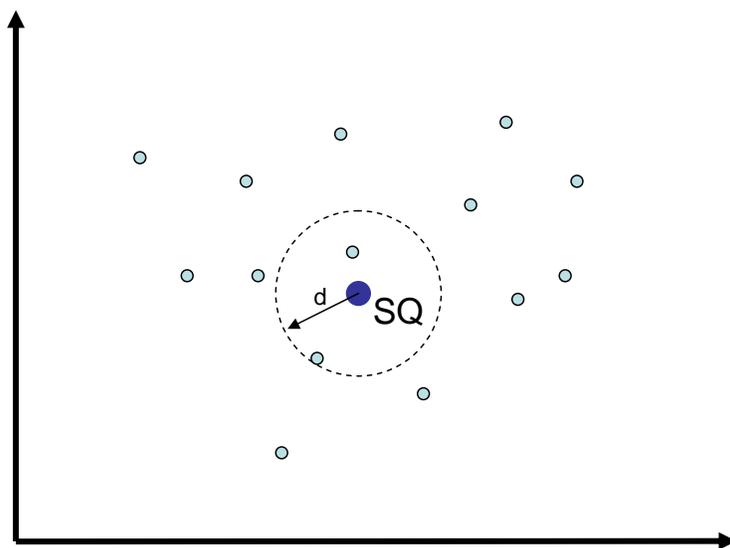
Figure 1: The basic game of international delegation. Smaller dots represent member states' ideal points while the large dot in the middle represents the status quo. The dotted circle indicates those outcomes that can be implemented under a rule limiting the organization's discretion to d.

events. If, for instance, all countries within the WTO play by the rules, then the policy outcome is that each country in the agreement receives Most-Favored Nation status. If one country imposes an import barrier in contravention to the WTO statutes, then by the organization's rules policy should change to one which punishes the offending state. For the moment it does not matter whether the change in policy effected by the organization's rules is implemented by a single individual, a body, or a completely mechanistic procedure of automatic adjustment.

Within this context, delegation is defined simply as the ability of the "entity" receiving authority to move policy away from the status quo. Along with the ability to move policy, though, we are interested in the degree to which policy can be changed — that is, the degree of *discretion* given to the international organization. We can envision limits on discretion as circumscribing the range of possible policies that the

entity can enact; for instance, a requirement that the policy outcome may not differ from the status quo by more than some distance d, as indicated in the figure. Then unbounded delegation is associated with infinitely large values of d, while as d shrinks to zero less and less power is delegated. When $d = 0$, delegation ends. We will define the degree of discretionary authority delegated to an international organization to be that organization's *power*.

Within this framework we can also define the *effectiveness* of an international organization as the difference between the final policy outcomes produced with the organization and the outcomes that would have resulted without it. If we find, for instance, that a given entity has no impact on final policy outcomes, then no matter how Byzantine its bureaucracy, how obscure its voting rules, how pious are countries' avowed intentions to adhere to its goals, we would be safe in labeling the organization as ineffective, and consequently would have little worry about any sovereignty costs associated with membership in that organization. And, purely as a definitional matter, an organization with no power is completely ineffective as well.

This view of delegation accords quite well with the terms of the debate laid out in the introductory essay by Bradley and Kelley (2007). For instance, it would concur that delegation does not include non-binding advisory authority, since pronouncements and resolutions do not on their own change the status quo policy. We believe that these powers are best thought of as creating focal points for interstate collective action, default actions that all may agree to follow at their will. This may well be important in certain cases — see the discussion below — but they do not in and of themselves change policy outcomes.

Our model also overlaps with Bradley and Kelley (2007) by not restricting the range of possible policies to those that would otherwise be exercised by the state.

It may, as the memo suggests, involve action in a truly international policy space, such as the regulation of international waters. Or it may involve the entity's control over the actions of *other* states in the organization, for instance by requiring them to impose import tariffs on goods exported by Country X. These actions certainly affect the welfare of Country X's citizens, and they certainly could not be implemented via the internal political processes of Country X alone.[2]

We also agree that authority must be delegated to an "entity," as long as it is clear that the entity in question may be nothing more than a set of predetermined policy responses, not an individual or group. But we think that perhaps it is not right to say that no delegation takes place if a country has veto authority over the decisions of the entity. This is, rather, a possible equilibrium outcome to a policy making game, and the entire game must be analyzed before concluding that a country's veto power protects it from delegations of sovereignty. Similarly, the memo does an excellent job of laying out a typography of different types of delegation and measures of their extent, but does not continue forward to enumerate the institutional limits on discretion that may accompany these delegations, such as the use of specific voting rules.

Within our setup, sovereignty costs can be defined as the distance between the policy that a country would implement if it were not a member of the international

---

[2]This raises, by the way, an interesting point about an organization's effectiveness and the presence of focal points or second-degree punishments. Say for instance that the US breaks a standard of international trade by levying anti-dumping duties inappropriately. In a world where there was only a norm of punishment for such behavior (leaving aside for the moment the problem of actually determining that the behavior was in fact inappropriate), the US might decide to take the action anyway and then engage in a series of bilateral discussions with the other countries to try and convince them not to retaliate. With the WTO in existence, though, it can issue a verdict instructing the other member countries to retaliate in a predetermined manner. Such an action, at the very least, creates a focal point for all other countries, making it easier for them to coordinate their responses. But the WTO goes further, making it *mandatory* within its rules to retaliate against the US. It is these second-degree punishments of countries that fail to carry out first-degree punishments that give the organization its real effectiveness.

organization, and the policy that it enacts once it has joined. In the figure, this would be the distance between the policy adopted by the international organization and each country's ideal point.

We can also incorporate the fact that some countries join an organization fully, while others do so only partially. In our model, this would occur when a country is allowed to move its policy only part of the distance away from its ideal point and towards the policy mandated by the international organization. For instance, one possibility in the two-dimensional space drawn here is that a country would have to adopt the organization's policy in the x-dimension, but not the y-dimension.

This brings us to the question of how the canonical delegation model should be modified for the international context. Two features stand out immediately: 1) that (some) member states often have veto power over the policy choices of the entity, and 2) that member states are free to drop out of the organization at any time. The first of these is also seen in certain domestic policy games — the legislative veto in the US, for instance. But the notion that any player can drop out of the organization is unique to the international context; the US House or Senate, for example, cannot choose to leave the polity just because they disagree with executive branch actions.

Taking these in turn, Figure 2 assumes that five member states are veto players (V). Consider status quos that lie within the area enclosed by the dotted lines joining the veto players' ideal points. It is clear that moving policy in any direction will make at least one veto player worse off, and thus in a single-shot game there would be no way to change a status quo like SQ. But a status quo outside of this region, such as SQ', could be changed, as long as it is moved closer to the veto players' desired policies. It is clear that the larger the area inside the dotted lines, the fewer the circumstances in which the organization can change policy away from the status
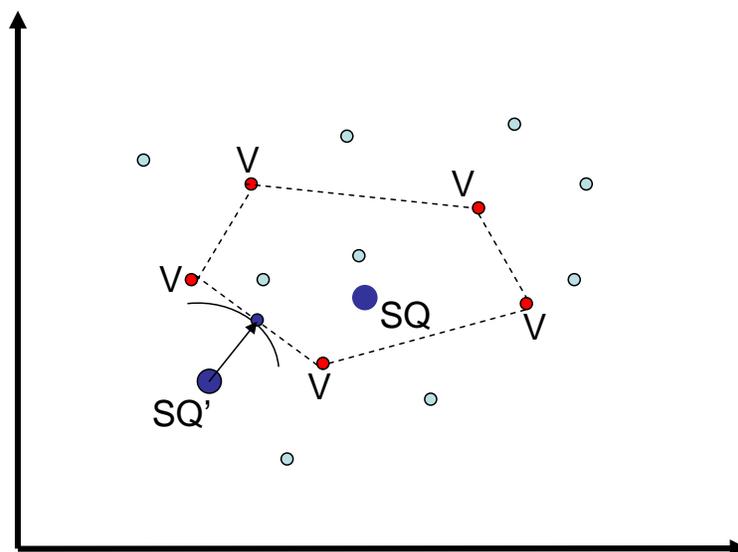
Figure 2: The delegation game with veto players.

quo, and hence the less effective the organization will be.

This model underlies the intuition that veto players in international organizations have not really delegated power. But there is more to the story. Organizations often engage in repeated interactions, not just single-shot games. In this context, a state with veto power might be convinced to allow a policy change that it does not agree with, in return for policy concessions further down the road. Insofar as the structure and content of these deals come from within the organization, real delegation is taking place, limited by the requirement that veto players are made better off in the long run by the sequence of decisions.

Furthermore, members of international organizations can leave any time they wish. We can assume that this would entail a loss of utility for the state leaving; otherwise, it would not have joined the organization. And it would probably make the other member states worse off as well; otherwise, they would not have allowed the

offending country into the organization in the first place.[3] This means that member states have a credible threat to exit the organization if continued adherence to its rules would decrease their utility too much.

Other states might well be willing to accept this exit option if the contemplated policy change would yield a utility increase that more than offset the loss of the state in question. Other times, though, the participation of a key state or small number of states may be so important that it is never optimal to have them exit. When this is the case, we would say that those countries *dominate* the organization.

Each dominant country acts as a one-way limit on the organization's discretion. Such a country is content to stay in the institution if the status quo policy is maintained — otherwise they would exit unilaterally — but if policy moves too far away from the status quo in a direction they are unhappy with, they would exit. And if multiple countries within the institution have similar preferences, the entire organization can become hamstrung and ineffective. Figure 3 illustrates one version of this scenario, in which three dominant countries leave essentially no room to deviate policy from the status quo without triggering a defection.

3   Model

We begin with a unidimensional policy space $X = [0, 1]$, where the value of x corresponds to the intensity of the regulatory regime in place. If X is the space of trade policy, for instance, then $x = 0$ would be an autarky with prohibitive tariffs, while $x = 1$ would be completely free trade. In nuclear policy, $x = 0$ might indicate no limits at all on nuclear development or testing, while $x = 1$ could mean an absolute

---

[3]One could imagine a situation, though, where this latter point does not hold. It may be, for instance, that one of the founding members of the organization is now a drag on everyone else's utility, but internal rules do not allow for its expulsion. This possibility would not qualitatively affect the results below.
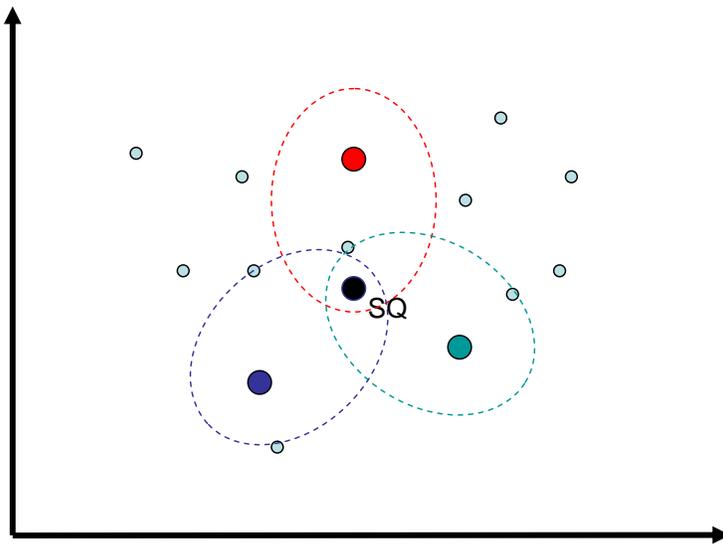
Figure 3: The delegation game with three dominant countries.

prohibition on such actions.

There is a set $\mathcal{N} = \{1, 2, \ldots, n\}$ of countries, each of which has domestic costs and benefits regarding the policy chosen. In particular, we assume that, acting in isolation from other countries, benefits per unit of x are constant, while the costs are increasing, so that for country i,

$$U_i(x) = x - \frac{1}{2\gamma_i}x^2.$$

Here, $\gamma_i \in [0, 1]$ is a parameter measuring the relative importance of costs to benefits for country i. Under our formulation, the relative costs of acceding to international demands rises at an increasing rate, relative to the benefits. So the political costs of free trade get higher the more domestic markets are opened, and eliminating the last 10 percent of pollutants that cause acid rain is more expensive than the first 10 percent. Solving for first order conditions shows that if only country i existed, it

would maximize utility by setting policy $x_i = \gamma_i$, making γ_i country i's "stand-alone" ideal point.

In addition to these domestic costs and benefits, each country can reap *associational benefits* by joining other countries in an international organization. Network externalities are captured by the assumption that each member state confers a benefit B_i on all other members. In trade policy, for example, B_i might be proportional to the size of a country's economy. In a mutual defence treaty, it could represent a country's military capacity. If a subset of countries $\mathcal{M} \subseteq \mathcal{N}$ join the organization as members, then potential associational benefits for each country are $B_{\mathcal{M}} = \sum_{i \in \mathcal{M}} B_i$. When $\mathcal{M} = \mathcal{N}$ we denote total benefits as simply B.

Member countries agree to set policy at $x_{\mathcal{M}}$, while non-members can set policy wherever they like. Associational benefits are then scaled by the level of policy, so that they are equal to $x_{\mathcal{M}} B_{\mathcal{M}}$ for all member countries. The idea here is that, for example, the benefits of free trade grow as the world trading organization requires higher and higher levels of tariff concessions. So if every country in the world joined the WTO, for instance, but $x_{\mathcal{M}} = 0$, then no free trade benefits would accrue in any case. Thus countries in the organization receive total utility

$$U_i(x, \mathcal{M}) = x_{\mathcal{M}} - \frac{1}{2\gamma_i} x_{\mathcal{M}}^2 + x_{\mathcal{M}} \sum_{j \in \mathcal{M}} B_j,$$

from remaining in the organization. This yields an adjusted ideal point $\gamma_i(1 + B_{\mathcal{M}}) = \gamma_i(1 + B)$ if all countries become members. So each country's "associational" ideal point $\gamma_i(1 + B)$ is larger — more internationally-oriented, that is — than their stand-alone ideal point.

Since membership in international organizations is voluntary, countries will only

remain in an organization as long as they receive at least as much utility as they would receive implementing their stand-alone ideal point; that is, as long as $U_i(x, \mathcal{M}) \geq U_i(\gamma_i)$. Such a country is willing to pay sovereignty costs equal to $|\gamma_i - x_{\mathcal{M}}|$, which is the distance from the policy imposed by the organization to the country's own ideal point. An organization's membership is globally stable when this condition holds for all $i \in \mathcal{N}$.

4 Results

From the preceding discussion, it is clear that absent any benefits from international cooperation, each country will set its own policy at its ideal point x_i. The greater the aggregate benefits B from joining the organization, the farther from its ideal point a given country is willing to allow policy to wander. So for a given international policy $x_{\mathcal{M}}$ and a given level of benefits B, country i is willing to join the organization as long as $x_{\mathcal{M}} \in [x_i - x_{\mathcal{M}}B, x_i + x_{\mathcal{M}}B]$.

Denote the range of policies that will induce country i to stay in the delegation regime $D_i(x_{\mathcal{M}}; B) = [x_i - x_{\mathcal{M}}B, x_i + x_{\mathcal{M}}B]$, which we will write as simply D_i when the association is clear. Then all countries will agree to stay in an international organization when these delegation ranges all overlap; that is, when

$$\bigcap_{i \in \mathcal{N}} D_i \neq \emptyset.$$

Order the countries so that $x_1 \leq x_2 \leq, \ldots, x_n$. It is clear that if $x_{\mathcal{M}} > x_1$, $x_{\mathcal{M}} \in [x_1 - x_{\mathcal{M}}B, x_1 + x_{\mathcal{M}}B]$, and $x_{\mathcal{M}} \in [x_n - x_{\mathcal{M}}B, x_n + x_{\mathcal{M}}B]$, then it must also be the case that $x_{\mathcal{M}} \in [x_j - x_{\mathcal{M}}B, x_j + x_{\mathcal{M}}B]$ for all $j \in 2, 3, \ldots, n - 1$. In other words, to check if a given delegation regime is globally stable we need only

check that countries 1 and n — those with extreme ideal points — prefer to stay in the organization rather than revert to their stand-alone ideal point. Without loss of generality, then, we can investigate questions of global stability assuming that $n = 2$.

To understand the equilibrium, consider a world where Country A prefers a low level of internationalization and Country B, a high level. For instance, a large industrialized nation prefers to make few concessions on reducing carbon emissions, while a smaller agrarian country would be happily emit few pollutants. Then the basic dilemma for the international organization is this: to provide any benefits from international standards it must raise the membership requirements $x_{\mathcal{M}}$ above zero, but if it raises them too far Country A might refuse to participate. The question is under what circumstances standards will exist which attract all countries to the agreement. And even if such a possibility does exist, will the organization prefer a relatively low level of standards with a higher membership, or a higher level that drives some low-demand countries away?

Formally, given two countries with ideal points x_1 and x_2 the total associational benefits associated with a policy level of $x_{\mathcal{M}}$ are $x_{\mathcal{M}}(B_1 + B_2) = x_{\mathcal{M}}B$. The overlap region $D_1 \cap D_2$ will be non-empty when $2x_{\mathcal{M}}B \geq x_2 - x_1$, in which case it will be equal to $D_{12}(x_{\mathcal{M}}) \equiv [x_2 - x_{\mathcal{M}}B, x_1 + x_{\mathcal{M}}B]$. The key to the equilibrium, as explained above, is that stability holds only when the policy generating this region is itself inside the region; that is, when $x_{\mathcal{M}} \in D_{12}(x_{\mathcal{M}})$.

When does this happen? The ranges D_1 and D_2 first touch at $\frac{x_1+x_2}{2}$. The $x_{\mathcal{M}}$

needed for this is:

$$
\begin{aligned}
x_{\mathcal{M}}B &= \frac{x_1 + x_2}{2} - x_1 \\
&= \frac{x_2 - x_1}{2} \\
x_{\mathcal{M}} &= \frac{x_2 - x_1}{2B}.
\end{aligned}
$$

This is to the left of the intersection when:

$$
\begin{aligned}
\frac{x_2 - x_1}{2B} &\leq \frac{x_1 + x_2}{2} \\
B &\geq \frac{x_2 - x_1}{x_2 + x_1} \equiv \underline{B}.
\end{aligned}
$$

Thus the grand coalition can be sustained whenever $B \geq \underline{B}$. Three propositions immediately follow: first, cooperation is easier when its benefits rise; no surprise there. Second, all else equal cooperation is easier when the countries have homogeneous preferences, so that $x_2 - x_1$ is small. And third, cooperation is easier when $x_1 + x_2$ is large, so that countries' stand-alone ideal points are more internationally oriented.

The lowest policy level which sustains cooperation occurs when

$$
\begin{aligned}
x_{\mathcal{M}} &= x_2 - x_{\mathcal{M}}B \\
x_{\mathcal{M}} &= \frac{x_2}{1 + B}.
\end{aligned}
$$

Similarly, the highest policy level which sustains cooperation occurs when

$$x_{\mathcal{M}} = x_1 + x_{\mathcal{M}}B$$

$$x_{\mathcal{M}} = \frac{x_1}{1 - B}.$$

This, in turn, is less than 1 when

$$B \leq 1 - x_1 \equiv \overline{B}.$$

We thus have three possibilities for the equilibrium:

1. $B < \underline{B}$: No level of $x_{\mathcal{M}}$ can induce global stability.

2. $B \in [\underline{B}, \overline{B}]$: global stability can be sustained for any $x_{\mathcal{M}} \in [\frac{x_2}{1+B}, \frac{x_1}{1-B}]$.

3. $B \geq \overline{B}$: global stability can be sustained for any $x_{\mathcal{M}} \in [\frac{x_2}{1+B}, 1]$.

In case 1 there is no way to sustain the grand coalition; the international organization cannot please all countries simultaneously. Therefore it will either not exist at all, but exist with less than a full complement of signatories. We consider this latter possibility more fully in the next section.

In case 3, the organization has a relatively free hand; it can set policy as high as it likes (up to $x_{\mathcal{M}} = 1$ and still retain full membership, as the benefits from association are high enough to induce even the low-demanders to stay in the organization.

In the middle case, cooperation is possible, but only for a limited range of values for $x_{\mathcal{M}}$. We have until now left the preferences of the international organization itself unspecified. Let us add a bit more structure by assuming two different types of organization. The first wishes to maximize the benefits of

internationalization, subject to the constraint of including as many countries as possible in the organization. The second simply wishes to maximize benefits and is willing to accept fewer members if necessary to accomplish these goals.

The latter type will be studied in the next section; for a membership-maximizing organization facing a "type two" equilibrium, the question is how far forward they can push $x_\mathcal{M}$ without losing members. The answer is $x_\mathcal{M} = \frac{x_1}{1-B}$, which is always greater than the midpoint between the ideal points x_1 and x_2. So on the one hand, outcomes in such organizations will be biased towards high-demanders; they pay fewer sovereignty costs than to the low-demanders. On the other hand, $\partial x_\mathcal{M}/\partial x_1 > 0$ while $\partial x_\mathcal{M}/\partial x_2 = 0$, so it is the low-demanders who have *power* within an organization; outcomes change in response to change in their preferences, not the high-demanders.

5 Extensions

Forthcoming...

6 Conclusion

In conclusion, one is tempted to say that delegating sovereignty is a bit of an oxymoron. Just like the U.S. Representative from Louisiana who claimed that his vote couldn't be bought, but it could be rented, we assert that sovereignty cannot be delegated, but it can be loaned. Nations participate in international organizations because it is in their interests to do so, and they can withdraw when this is no longer the case. This does not mean that no alternative arrangement could make the country better off, or that their preferred policies are always enacted. Barring coercion or extreme exit costs, though, participation in international organizations

is voluntary and therefore should be seen as a natural extension of member states' rights and an exercise of their sovereignty.

The problem, we argued, is that formal models had not yet been developed in which sovereignty costs and associational benefits could be addressed in a natural manner. We attempted to rectify this situation by describing a basic model built on the twin assumptions of free association and network externalities. We found that equilibrium conditions were straightforward to characterize and that, in general, low-demand countries pay higher sovereignty costs but wield more power in international organizations than do their low-demand counterparts.