SOVEREIGNTY AND DELEGATION IN INTERNATIONAL ORGANIZATIONS

DAVID EPSTEIN*

SHARYN O'HALLORAN**

I

INTRODUCTION

Established in 1945, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) is a specialized agency of the United Nations established to "contribute to peace and security by promoting international collaboration through education, science, and culture."[1] After a peaceful beginning, UNESCO became embroiled in controversy with the 1980 publication of the MacBride Report, which called for the democratization of communication and strengthening of national media.[2] The United States and the United Kingdom denounced the report as an attack on freedom of the press and criticized the organization in general as a platform for communist and Third World countries to attack the West.[3]

The United States withdrew its funding for and membership in UNESCO in 1984, followed by the United Kingdom the next year.[4] In the ensuing decade, UNESCO toned down its rhetoric and reorganized itself to be less top-heavy. The United Kingdom eventually rejoined UNESCO in 1997,[5] and the United States rejoined in 2003.[6] How can we explain the exit and reentry of these two crucial countries? Why were they willing to be associated with UNESCO for

* Professor of Political Science, Columbia University.

** George Blumenthal Professor and Professor of Political Science and International and Public Affairs at Columbia University.

1. Global Embassy, Goodwill Ambassadors UNESCO, http://www.globalembassy.org/goodwillambass_UNESCO.htm (last visited Sept. 5, 2007).

2. *See* Amit Mukherjee, *International Protection of Journalists: Problem, Practice, and Prospects*, 11 ARIZ. J. INT'L & COMP. LAW 339, 349 (1994).

3. *See* Sean D. Murphy, *Contemporary Practice of the United States Relating to International Law: International Organization: United States' Return to UNESCO*, 97 AM. J. INT'L L. 977, 977–78 (2003).

4. Anna Gercas, *The Universal Declaration on Bioethics and Human Rights: Promoting International Discussion on the Morality of Non-Therapeutic Research on Children*, 27 MICH. J. INT'L L. 629, 632 n.11 (2006).

5. The White House, Fact Sheet: United States Rejoins UNESCO, http://www.whitehouse.gov/news/releases/2002/09/20020912-4.html (last visited Sept. 5, 2007).

6. U.S. Dep't of State, United States of America and UNESCO: Building Knowledge, Bridging Culture, http://usinfo.state.gov/products/pubs/unesco/ (last visited Sept. 5, 2007).

one period of time but not another, and how did their behavior affect the organization's policies?

The UNESCO case highlights a key feature of delegation relationships within international organizations (IOs)—namely, that membership in such organizations is voluntary. What sets IOs apart from countries' internal delegation regimes is the fact that if a country is not satisfied with the results it is obtaining via membership in the organization, it can simply exit, as the United States and the United Kingdom did from UNESCO, or simply decline to join the organization in the first place.[7] This stands in contrast to, for instance, interbranch delegation; if Congress is unhappy with the executive's use of delegated authority, it cannot simply leave and declare itself to be the national legislature of Bolivia instead. More to the point, any change in the delegation regime would itself be subject to a presidential veto, so each Congress is to some degree locked into the delegation arrangements inherited from previous Congresses.

If IO membership is voluntary, why would countries delegate in the first place? We argue that international organizations are held together by network externalities, such as free trade, safety via nuclear nonproliferation, and so on. Specifically, a defining feature of international organizations is that the more countries that belong to them, the more benefits accrue to all members. In this sense, IOs display increasing returns to scale, similar to many social or Internet-based resources. Conversely, the departure of key countries can do significant harm to an international organization, sometimes triggering a wave of defections.[8]

This article provides a theory of delegation to IOs that incorporates free exit and network externalities into the standard delegation-modeling framework. What issues should such a theory to be able to address?

1. It should predict an IO's membership, including states' decisions to enter and exit.
2. It should predict the policy goals pursued by the IO; moreover, these policy goals should themselves affect membership. Formal models of IOs to date take either the policy choices or the member states as given, but the UNESCO example above makes clear that changes in policy can lead to changes in membership as well.

---

7. Whether exit is truly costless is open to debate; some international organizations place limits on the rights of their members to exit, although the degree to which these actually bind member states is unclear. For a full discussion, see generally Lawrence Helfer, *Exiting Treaties*, 91 VA. L. REV. 1579 (2005). In any case, this article's argument relates international conditions to an IO's membership, whether through previous entry and exit, or whether a country never joins the IO at all, as the United States did with the League of Nations.

8. In fact, the United States' departure from UNESCO was followed not only by that of the United Kingdom, but also by that of Singapore in 1986. UNESCO EXECUTIVE BOARD, REPORT BY THE DIRECTOR-GENERAL ON BUDGET ADJUSTMENTS 6 (1997), *available at* http://unesdoc.unesco.org/images/0010/001087/108711E.pdf.

3. It should allow for differences among states; in particular, some states might be in a position to confer more benefits on their fellow member states than others, such as large countries offering access to their markets by lowering trade barriers. These differences are, in turn, the source of differential power among member countries within the IO.
4. Consistent with the themes of this symposium, it should logically define and incorporate the notions of delegation and sovereignty costs. Even in the face of these costs, it should give countries an incentive to join and stay in the IO.
5. Finally, it should allow for the possibility of partial accession to treaties and predict under which circumstances devices such as reservations will be allowed.

The model we present below is able to address all these issues, so it is a useful prototype to study delegation relationships within IOs. In contrast to the standard setting, where the degree of discretion delegated is linked to the harmony of interests among the actors, in the international setting an IO's authority will be circumscribed by the "biggest weakest link;" that is, the least internationally oriented country whose exit would cause the IO to collapse.

The next Part reviews the relevant modeling literature on delegation in IOs. Then, in the third Part, we provide an overview of our approach and relate it to the terms of debate in the sovereignty literature. The Part that follows provides a formal specification of the model. The fifth Part solves for the equilibria of the model in general, while the sixth provides a detailed examination of a three-country case with an integrationist IO. The last Part concludes with implications for the long run sustainability and optimal structure of international organizations.

II

LITERATURE REVIEW

There are many excellent studies of international organizations that emphasize the delegation relationships inherent in their establishment and operation. Daniel Nelson and Michael Tierney, for instance, have reviewed agency theory and applied it to World Bank environmental reform projects.[9] Other essays explore the ways in which countries circumscribe delegation relationships with IOs so as to limit agency losses.[10]

There are, however, few formal models of delegation to IOs; rather, game-theoretic treatments of IOs tend to explore IOs' relation to the collective dilemmas that bring them into existence. The modeling tradition begun by Robert Keohane, for example, views international organizations as solutions to

9. *See generally* Daniel Nelson & Michael Tierney, *Delegation to International Organizations: Agency Theory and World Bank Environmental Reform,* 57 INT'L ORG. 241 (2003).

10. *See, e.g.,* DARREN HAWKINS ET AL., DELEGATION AND AGENCY IN INTERNATIONAL ORGANIZATIONS (2006).

coordination problems that countries cannot solve efficiently through a series of bilateral agreements.[11] A second line of reasoning, due originally to George Tsebelis, examines the impact of domestic bargaining on the nested or "two-level" games of international relations.[12] And recently, a third line of inquiry has emphasized the informational advantages conferred by membership in international organizations.[13]

But given their description of international cooperation as simply a solution to a prisoners' dilemma, these models cannot easily incorporate concepts such as sovereignty costs, partial association or, indeed, power in international organizations. Nor do they simultaneously address the questions of which policies IOs adopt and which countries choose to belong to IOs, let alone the more complicated issues of reservations and reciprocity. This article contributes to the literature precisely along these lines of inquiry.

The model has as its basis the classic models of delegation relationships in political science.[14] It is closely associated with the articles by Mathew McCubbins, Roger Noll, and Barry Weingast and by us,[15] in which discretion was essentially defined as "whatever the agency can get away with," rather than being explicitly circumscribed by the political actors delegating to the agency. But rather than require all other political players to act in concert to overturn an agency's decision—such as when the House, Senate, and President must each approve a new law to counteract agency rulemaking—the setting for this model allows any single member country to unilaterally change outcomes by simply removing itself from the IO.

III

MODEL PREVIEW

The theoretical framework for analyzing delegation in political science is a variant on the spatial model of political institutions, which assumes that actors have most-preferred policies—or ideal points—in some policy space, and they want to bring policy outcomes as close as possible to their ideal point. It is political institutions that determine who gets enfranchised into the

11. *See generally* ROBERT O. KEOHANE, AFTER HEMOGENY: COOPERATION AND DISCORD IN THE WORLD POLITICAL ECONOMY (1984).

12. *See generally, e.g.,* GEORGE TSEBELIS, NESTED GAMES: RATIONAL CHOICE IN COMPARATIVE POLITICS (1991).

13. *See, e.g.,* Leslie Johns, *A Servant of Two Masters: Communication and the Selection of International Bureaucrats*, 61 INT'L ORG. 245 (2007); Andrew Kydd, *Which Side Are You On? Bias, Credibility, and Mediation*, 47(4) AM. J. OF POL. SCI. 597 (2003); Helen Milner & B. Peter Rosendorff, *Trade Negotiations, Information and Domestic Politics: The Role of Domestic Groups* 8(2) ECON. & POL. 145–89 (1996).

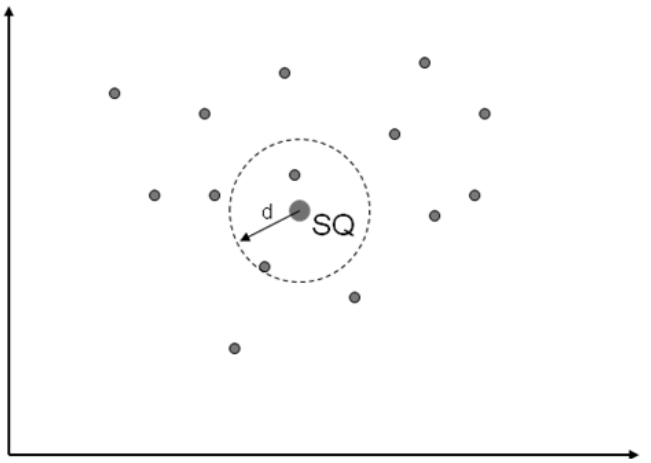
14. *See, e.g., infra* notes 15–16.

15. *See, e.g.*, DAVID EPSTEIN & SHARYN O'HALLORAN, DELEGATING POWERS (1999) [hereinafter EPSTEIN & O'HALLORAN, DELEGATING POWERS]; David Epstein & Sharyn O'Halloran, *Administrative Procedures, Information, and Agency Discretion,* 38(3) AM. J. POL. SCI. 697 (1994) [hereinafter Epstein & O'Halloran, *Administrative Procedures*]; M.D. McCubbins et al., *Administrative Procedures as Instruments of Political Control*, 3 J. L. ECON. & ORG. 243 (1987).

policymaking process, under what conditions, and thus who wins and who loses at the end of the day.

This setup is illustrated in two dimensions in Figure 1. The small dots represent the ideal points of member states in some international organization. The large dot indicates the status quo policy ex ante. The dotted circle indicates those outcomes that can be implemented under a rule limiting the organization's discretion to *d*.

Figure 1 : The basic game of international delegation.



The policymaking institutions of the organization determine which way policy will move in response to various external events. If, for instance, all countries within the World Trade Organization (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 of the WTO statutes, then by the organization's rules the offending state should be punished. 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 conferring to the IO some degree of authority to move policy away from the status quo.[16] Of particular interest is the distance from the status quo 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 IO 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

16. In the domestic-politics setting, an agency can usually move policy unilaterally. In the international setting with free exit, the IO must induce each member state to change its policies, if necessary, to be brought into line with its dictates.

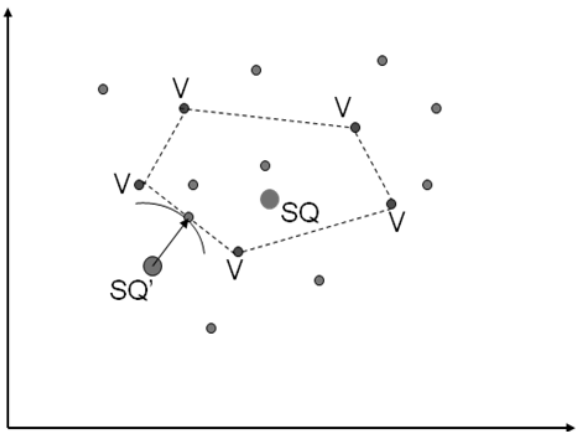
infinitely large values of *d*; as *d* shrinks towards zero, less and less power is delegated. When $d = 0$, delegation ends.

This setup has the advantage of making certain terms in the delegation debate susceptible to precise definition. Most importantly, 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 organization and of 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. Therefore, sovereignty costs are no more or less than changes in policy outcomes.

The model can also incorporate the fact that some countries join an organization fully, while others do so only partially. In the 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. Finally, this framework can clarify some questions in the IO literature, such as whether countries with veto power actually delegate.

Figure 2, for example, assumes that five member states are veto players (V).

Figure 2: The delegation game with veto players.



Consider status quo points that lie within the area enclosed by the dotted lines joining the veto players' ideal points. 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. The larger the area inside the dotted lines—that is, the more heterogeneous the preferences of the veto players—the fewer the circumstances in which the organization can change policy away from the status quo, and hence the less discretion the organization will have. This finding that

organizations have more discretion when the preferences of their member states are more homogeneous could explain, for instance, why many European Union nations resisted expansion, despite the network benefits to be gained by adding more countries.

Since there are circumstances in which the IO can move policy, being a veto player does *not* nullify delegation. On the other hand, since all changes in policy must make each veto player at least indifferent between the proposed policy and the status quo, veto players will not pay sovereignty costs by joining the IO.[17]

IV

MODEL

A variant on the traditional delegation model captures the dual features of free exit and network externalities in international organizations. Although this variant borrows the intuition from the two-dimensional model discussed in the previous Part, here a uni-dimensional policy space, $X = [0,1]$, is investigated, where the value of x corresponds to the amount of the public good or cooperative behavior provided. 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 proliferation, $x = 0$ might indicate no limits at all on nuclear development or testing, while $x = 1$ could mean an absolute prohibition on all 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 political importance of regulatory costs for Country i. This formulation emphasizes the fact that, to varying degrees, countries will often provide some amount of international public goods on their own; even absent trade agreements, for example, few countries are run as complete autarkies.

The relative costs of acceding to international demands rise 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 ten percent of

17. Furthermore, member states of IOs 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. That is, individual decisions may impose sovereignty costs even on a veto player if the long-run tradeoffs are important enough.

pollutants that cause acid rain is more expensive than eliminating the first ten percent. Solving for first-order conditions shows that if only Country *i* existed, it would maximize utility by setting policy $x_i = \gamma_i$, where x_i is 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 of an IO. In trade policy, for example, b_i might be proportional to the size of the tradable sector of a country's economy. In a mutual defense 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 $\mathcal{M} = \mathcal{N}$, total benefits are denoted as B.

In the game, the IO moves first, naming a policy level x. All countries then simultaneously and noncooperatively decide whether to become a member of the IO. Member countries agree to set policy at x, while nonmembers can set policy wherever they like. Associational benefits are then scaled by the level of policy, so that they are equal to $xB_{\mathcal{M}}$ for all member countries. The idea here is that, for example, the benefits of free trade grow as the WTO requires higher and higher levels of tariff concessions. So if every country in the world joined the WTO, for instance, but $x = 0$, then no free-trade benefits would accrue in any case. Thus countries in the organization receive total utility:

$$\begin{aligned} \mathbf{U}_i(x, \mathcal{M}) &= U_i(x) + xB_{\mathcal{M}} \\ &= x - \frac{1}{2\gamma_i}x^2 + x \sum_{j \in \mathcal{M}} b_j \end{aligned}$$

from remaining in the organization. This yields an adjusted ideal point $x_i(1 + B_{\mathcal{M}}) = x_i(1 + B)$ if all countries become members. Each country's "associational" ideal point $x_i(1 + B)$ is larger—more internationally oriented, that is—than their stand-alone ideal point.[18]

Since membership in international organizations is voluntary, countries will remain in an organization only 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

$$\mathbf{U}_i(x, \mathcal{M}) \geq U_i(x_i).$$

18. This is the parallel in our model to Robert Keohane's assumption that all countries would agree that they would be made better off if they could all be forced to contribute more to the public good than they each would contribute individually. *See* ROBERT KEOHANE, *supra* note 11, at 67–78.

Such a country is willing to pay sovereignty costs equal to $|x_i - x|$, 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 $i \in \mathcal{N}$.

V

DELEGATION AND FEASIBLE COALITIONS

Assume that the IO wishes to attract member states $\mathcal{M} \subseteq \mathcal{N}$. This Part derives conditions under which such a coalition can form, and if so, which policies the IO can enact. From the preceding discussion, it is clear that absent any benefits from international cooperation, each country will set policy at its ideal point x_i. Furthermore, for any value of x, there is an equilibrium where no state joins the IO, since an organization with no members cannot induce any one state to alter its policy away from its ideal point. Similarly, changing policy so that a particular country decides to drop out might trigger a cascade of defections, since the aggregate benefits of remaining in the organization have been reduced. The analysis will thus focus on the equilibrium with the largest possible sustainable membership in the IO.

The greater the aggregate benefits $B_{\mathcal{M}}$ from joining the organization, the farther from its ideal point a given country is willing to allow policy to wander. So for a given policy *x*, potential member countries $\mathcal{M} = \{1, 2, \ldots, m\}$, and associated level of benefits $B_{\mathcal{M}}$, Country i is willing to join the organization as long as $x \in [x_i - xB_{\mathcal{M}}, x_i + xB_{\mathcal{M}}] \equiv D_i$, Country *i*'s "delegation range." Then all countries in $\mathcal{M}$ will agree to stay in the IO when these delegation ranges overlap; that is, when

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

Order the countries in $\mathcal{M}$ so that $x_1 \leq x_2 \leq, \ldots, x_m$, so that Country 1 is the low-demander and Country *m* is the high-demander. If $x \in D_1$, and $x \in D_m$, then it is easy to see that $x \in D_j$ for all $j \in 2, 3, \ldots, m-1$. In other words, to check if a given delegation regime is stable for all countries in $\mathcal{M}$, one need only check that Countries 1 and m—those with extreme ideal points—prefer to stay in the organization rather than revert to their stand-alone ideal policies.

Formally, for a given policy level *x*, the overlap region $D_1 \cap D_m$ will be nonempty when $2xB_{\mathcal{M}} \geq x_m - x_1$, in which case it will be equal to $D_{1m}(x) \equiv [x_m - xB_{\mathcal{M}}, x_1 + xB_{\mathcal{M}}]$. The key to the equilibrium is that the policy *x* generating this interval must also be inside it: we must find some *x* for which $x \in D_{1m}(x)$. Otherwise, the degree of internationalization necessary to generate

sufficient network externalities will ask too much of the low-demanding country.[19]

When will the delegation ranges intersect in the appropriate way? The ranges D_1 and D_m first touch at $\frac{x_1+x_m}{2}$. The x needed for this is

$$\begin{aligned} xB_{\mathcal{M}} &= \frac{x_1+x_m}{2} - x_1 \\ &= \frac{x_m - x_1}{2} \\ x &= \frac{x_m - x_1}{2B_{\mathcal{M}}}. \end{aligned}$$

This is to the left of the intersection when:

$$\begin{aligned} \frac{x_m - x_1}{2B_{\mathcal{M}}} &\leq \frac{x_1+x_m}{2} \\ B_{\mathcal{M}} &\geq \frac{x_m - x_1}{x_m + x_1} \equiv \underline{B}(\mathcal{M}), \text{(Equation 5.1)} \end{aligned}$$

where $\underline{B}(\mathcal{M})$ denotes the minimum level of benefits B that can support a coalition with member states $\mathcal{M}$.

Thus coalition $\mathcal{M}$ can be sustained whenever $B_{\mathcal{M}} \geq \frac{x_m - x_1}{x_m + x_1}$. Three propositions immediately follow: First, cooperation is easier when its benefits rise; no surprise here. Second, all else being equal, cooperation is easier when the countries have homogeneous preferences, so that $x_m - x_1$ is small. And third, cooperation is easier when $x_1 + x_m$ is large, so that countries' stand-alone ideal points are more internationally oriented.

The lowest policy level which sustains cooperation occurs when

$$\begin{aligned} x &= x_m - xB_{\mathcal{M}} \\ x &= \frac{x_m}{1 + B_{\mathcal{M}}}. \end{aligned}$$

19. For example, the proposed United States–South Korea free-trade agreement is currently foundering over the issue of imported beef. The South Korean government would, in principle, like greater integration, but the political costs sustained by exposing the beef industry to competition would overwhelm any compensating benefits, and without such a provision the agreement would be too weak to attract U.S. participation. For the U.S. position on the issue, see Embassy of the United States, U.S. Free-Trade Pact with South Korea Would Enhance Partnership (June 13, 2007), http://seoul.usembassy.gov/413_061407a.html (last visited Jan. 29, 2008).

Similarly, the highest policy level which sustains cooperation occurs when

$$\begin{aligned} x &= x_1 + xB_{\mathcal{M}} \\ x &= \frac{x_1}{1 - B_{\mathcal{M}}}. \end{aligned}$$

This, in turn, is less than 1 when

$$B_{\mathcal{M}} \leq 1 - x_1 \equiv \overline{B}(\mathcal{M}), \text{(Equation 5.2)}$$

where $\overline{B}(\mathcal{M})$ denotes the level of B above which all member states in coalition $\mathcal{M}$ will remain in the IO, even with $x = 1$.

We thus have three possibilities for the equilibrium:

1. $B_{\mathcal{M}} < \underline{B}(\mathcal{M})$: Coalition $\mathcal{M}$ cannot form.
2. $B_{\mathcal{M}} \in [\underline{B}(\mathcal{M}), \overline{B}(\mathcal{M})]$: Coalition $\mathcal{M}$ can be sustained for any $x \in [\frac{x_m}{1+B_{\mathcal{M}}}, \frac{x_1}{1-B_{\mathcal{M}}}]$.
3. $B_{\mathcal{M}} > \overline{B}(\mathcal{M})$: Coalition $\mathcal{M}$ can be sustained for any $x \in [\frac{x_m}{1+xB_{\mathcal{M}}}, 1]$.

In the first case, there is no way to sustain the proposed coalition $\mathcal{M}$; the international organization cannot please all countries simultaneously. In particular, policies that would induce Country 1 to join the IO have too little spillover benefits to attract Country *m*, while policies that would keep Country *m* satisfied are too demanding for Country 1. In the third case, the organization has a relatively free hand; it can set policy as high as it likes (up to $x = 1$) and still retain all states in the coalition, because the benefits from association are high enough to induce even the low-demanders to stay in the organization.

In the second case, cooperation is possible, but only for a limited range of values for x. Assuming that the IO implements the highest degree of integration possible, it will set $x = \frac{x_1}{1-B_{\mathcal{M}}}$, which is always greater than the midpoint between the ideal points x_1 and x_m. So on the one hand, outcomes in such organizations will be biased towards high-demanders; they pay fewer sovereignty costs than do the low-demanders. On the other hand, $\partial x/\partial x_1 > 0$ while $\partial x/\partial x_m = 0$, so it is the low-demanders who have *power* within an organization; outcomes change in response to changes in their preferences, not those of the high-demanders. The limits of an IO's discretion are therefore determined by its "biggest weakest link," or the lowest-demand country that the IO wishes to include in the organization.

VI

EQUILIBRIUM WITH AN INTEGRATIONIST ORGANIZATION

Part V detailed the conditions under which an equilibrium including all states in coalition $\mathcal{M} \subseteq \mathcal{N}$ could exist, and if existing, the ranges of feasible policies that the IO could implement. It also detailed the basic tradeoff inherent in determining an IO's membership: expanding the list of participating countries versus higher levels of integration. The next step is to derive predictions as to which of the feasible coalitions will actually constitute the IO, and to do this one must first specify the IO's preferences over outcomes, for both the policy chosen and the extent of its membership.

Analogous to the traditional delegation literature, one could assume that the IO has an ideal point within the policy space and tries to bring outcomes as close to that point as possible. That is, the IO could be modeled as if it were a state within the system, and this might be a reasonable approach for organizations that are dominated by a single country (as, arguably, the World Bank is dominated by the United States). Similarly, if policies are to be voted on, the IO might adopt the ideal point of its median member, or the pivotal member state in a supermajority or qualified majority voting system. Alternatively, one might assume that organizations have position-taking preferences—much as agencies are sometimes modeled as wanting to maximize their budgets—in which case they might take extreme stances on issues important to those who run the IO, regardless of the impact this might have on the organization's reputation or membership. (This is one interpretation of UNESCO's actions as detailed in Part I.) Another possibility is that the IO cares only about maximizing membership, regardless of the policy consequences. The United Nations, for instance, feels that it can only operate legitimately if it can boast global membership, even though this means that it finds itself hamstrung in many cases, unable to act without triggering a veto by key member states.

Again, though, the translation of delegation models to an international setting provides extra possibilities. In particular, we assume that the IO has what we term "integrationist preferences," meaning that it wants to maximize $xB_{\mathcal{M}}$, the weighted sum of the policy positions for states in the IO. For instance, the WTO would be modeled as wanting to maximize the amount of free trade in the world trading system. Such a formulation means that the IO will take into account the fact that raising standards x will make trade freer for those countries that continue to participate in the system, but it might reduce free trade overall by inducing some states to exit. Thus, integrationist organizations will under some circumstances take into account the preferences of those countries less oriented towards integrationist policies, especially if those countries weigh heavily in terms of the potential benefits they offer (high values of b_i), or if their leaving the system could induce a cascade of defections by other countries.

We examine a version of this game with three countries, having ideal points $0 < x_1 < x_2 < x_3 \leq 1$. To simplify the analysis, we assume the countries are identical except for their ideal points, so $b_1 = b_2 = b_3 = b$. Since a nontrivial IO has more than one member, there are three possible equilibrium configurations: no countries join the IO (it cannot form a coalition around any policy); Countries 2 and 3 join the IO; or all three countries join.

Equations 5.1 and 5.2 above provide the conditions under which either the coalition of Countries 2 and 3 or the grand coalition (GC) of all countries can be maintained:

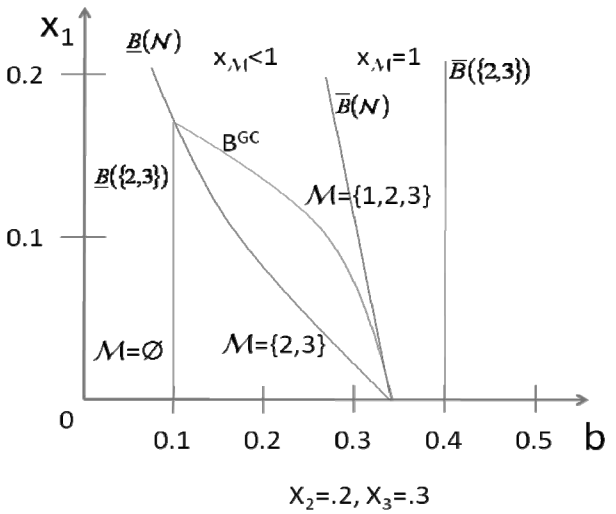
$$\begin{aligned} \underline{B}(2,3) &= \tfrac{x_3 - x_2}{2(x_3 + x_2)} \\ \overline{B}(2,3) &= \tfrac{1 - x_2}{2} \\ \underline{B}(\mathcal{N}) &= \tfrac{x_3 - x_1}{3(x_3 + x_1)} \\ \overline{B}(\mathcal{N}) &= \tfrac{1 - x_1}{3}. \end{aligned}$$

If $b < \min\left(\underline{B}(2,3), \underline{B}(\mathcal{N})\right)$, then no IO can form, as network externalities are too small to entice countries away from their stand-alone ideal points. It remains to specify the conditions under which the IO will choose to set policy to attract only countries 2 and 3, and when it will attract all three to a grand coalition. The latter provides the IO with greater utility when

$$\begin{aligned} 3b\frac{x_1}{1 - 3b} &\geq 2b\frac{x_2}{1 - 2b} \\ b &\geq \frac{2x_2 - 3x_1}{6(x_2 - x_1)} \equiv B^{GC}. \end{aligned}$$

Equilibrium outcomes when $x_2 = 0.2$ and $x_3 = 0.3$ are shown in Figure 3 for different values of network benefits b and Country 1's ideal point x_1. For any given level of x_1, $\underline{B}(2,3)$ and $\underline{B}(\mathcal{N})$ give the minimum values of b needed to sustain cooperation. For points where some cooperation is possible, the B^{GC} line separates those cases where an integrationist IO prefers the grand coalition to attracting only Countries 2 and 3.

Figure 3: Three-country equilibrium when IO has integrationist preferences. The Figure assumes that $x_2 = 0.2$ and $x_3 = 0.3$.



As indicated, in equilibrium, low values of b are associated with no organization at all ($\mathcal{M} = \emptyset$). High values of both x_1 and b have the grand coalition ($\mathcal{M} = \{1, 2, 3\}$), and intermediate values have a coalition of Countries 2 and 3 ($\mathcal{M} = \{2, 3\}$). Since all points in the area showing coalitions of Countries 2 and 3 lie below $\overline{B}(2, 3)$, they all involve levels of cooperation with $x < 1$. For points above B^{GC} but to the left of $\overline{B}(\mathcal{N})$, x is again less than 1, while to the right of $\overline{B}(\mathcal{N})$, maximum cooperation with a grand coalition is possible.

What does this diagram say about the membership and goals of IOs? First, they are most effective when preferences are homogeneous, so that Country 1 is less of an outlier, and when the potential benefits of cooperation are high. In those areas where the grand coalition forms, though, Country 1 pays fewer sovereignty costs and has more power within the IO when it is more of an outlier, because the IO will prefer to accommodate Country 1's demands and will set policy in a manner that is sensitive to changes in Country 1's ideal point. But unless the benefits from cooperation are high enough, there will come a point at which Country 1 is too much of an outlier and the organization will allow it to exit (or not give it sufficient incentives to join in the first place), thus eliminating Country 1's sovereignty costs altogether but, of course, also eliminating any power or leverage that Country 1 had with the organization.

This type of framework facilitates certain thought experiments with respect to specific incidents in the development of IOs. For instance, at first blush the interpretation of the UNESCO case cited in Part I is that the IO initially had

extreme position-taking preferences, so that it cared more about making a statement than retaining the support of certain major countries. If this view is correct, then one would look for a leadership change within UNESCO, possibly chastened by the exits of several countries, which would lead to new policies sufficiently reducing the sovereignty costs so that the United States and United Kingdom would return to the fold.

But there are other possibilities as well. Say the ideal points of the United States and United Kingdom were to become more extreme relative to the other countries already in UNESCO—either through their becoming more conservative or through other member countries becoming more internationally oriented—so that an integrationist organization preferred to move away from the grand coalition. Then one would look to changes in the preferences of member countries to explain the reintegration of the defectors. Or say the benefits from cooperation were to fall so that it was no longer worthwhile to keep the United States and the United Kingdom within the IO. In that case, one would look to changes in technology or other elements of the international policy environment to explain reentry. Or perhaps it would be some combination of the above. None of these perspectives is necessarily the correct one; the point is that having a modeling framework suggests checking different possibilities and empirical correlates than would have occurred to researchers beforehand.

VII

CONCLUSION

One is tempted to say that delegating sovereignty is a bit of an oxymoron. Just like the U.S. Senator from Louisiana who insisted that his vote could not be bought, but it could be rented, in a world with free exit, sovereignty cannot truly be delegated, but it can be loaned. Nations participate in international organizations because it is in their interest to do so, and they can withdraw when this is no longer the case. This does not mean that no alternative arrangement might 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, rather than as a violation.

Our intent in this article is to offer an extension of formal models of delegation that takes free exit and network externalities into account, and in which sovereignty costs, IO membership, and international policy can be addressed in a natural manner. The model shows that equilibrium conditions are straightforward to characterize and that, in general, low-demand outlier countries pay higher sovereignty costs but wield more power in international organizations than do their more internationally-oriented counterparts.

This model can be extended in a number of interesting ways. First, partial adoption of treaties can be integrated into the analysis by allowing member countries to adopt any policy $x = \beta_i x$, where $\beta_i \in [0, 1]$ is a parameter indicating the proportion of the IO's mandates adopted by Country i. Such a country would then contribute network externalities equal to $\beta_i b_i$ toward the common good and, if other countries reciprocate, receive only $\beta_i B_{\mathcal{N}}$ benefits in return. One could then solve for the optimal β_i for each country and see when an integrationist IO would allow such reservations as the necessary price for retaining the membership of certain key countries. We leave these and other extensions to future work.