

## The Influence of International Institutions

INSTITUTIONAL DESIGN, COMPLIANCE, EFFECTIVENESS,  
AND ENDOGENEITY

*Ronald B. Mitchell*

IDENTIFYING THE INFLUENCE of international institutions requires taking into account how problem structure influences institutional design. Three claims are central to the relationships among international institutions, state interests, and state behaviors. First, states act to further their goals; second, states use international institutions to advance those goals and “design institutions accordingly”; and third, states “fight over institutional design because it affects outcomes.”<sup>1</sup> Yet if states behave in ways that reflect their goals and interests but also push for designs of international institutions that reflect those goals and interests, how can those institutions be said to “affect outcomes” in the sense of causing states to behave differently than they would have otherwise? Making sense of these claims simultaneously requires us to assume (and have a logic for assuming) that the goals and interests that states codify in international institutions differ from the goals and interests that drive their behavior.

In short, to demonstrate the influence of international institutions on state behavior convincingly requires taking the realist challenge seriously. The ability of international institutions to promote cooperation in the presence of preexisting conflict is central to Robert Keohane’s intellectual legacy.<sup>2</sup> Realists contest this claim about institutions, arguing that states’ interests drive not only their behavior, but also both the form of the institutions they create and their choices about membership in them. Therefore, those institutions cannot have any independent effect on their behavior. Behavior is driven either directly by structure or indirectly by

Acknowledgments: This chapter has benefited from valuable comments and discussions with R. Charli Carpenter, Page Fortna, Erin Graham, Helen Milner, and Craig Parsons. The material presented here is based upon work supported by the National Science Foundation under Grant No. 0318374 entitled “Analysis of the Effects of Environmental Treaties,” September 2003–August 2008. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

<sup>1</sup> Koremenos, Lipson, and Snidal 2001, 762.

<sup>2</sup> Keohane 1984.

“structure-through-institutions” with any observed institutional “effects” being spurious or epiphenomenal.<sup>3</sup>

The realist challenge involves two different charges of endogeneity. *Membership endogeneity* involves the claim that, for any international institution, countries that join differ systematically from those that do not. Those preexisting differences drive the decision to join and explain any subsequent differences in behavior between members and nonmembers. What appear to be institutional influences are simply the postagreement expression of preexisting differences. *Design endogeneity* involves the claim that variation in institutional design reflects systematic differences in the underlying structure of the problem being addressed. In this case, differences in problem structure dictate the variation in what provisions countries negotiate and accept in international agreements and also explain systematic differences in how states respond to those agreements. What appears as one institution wielding greater influence than another is simply the postagreement expression of differences in problem structure. Thus, an individual institution's effects cannot be identified by contrasting the behaviors of members to nonmembers because countries are not assigned randomly to treaties. Nor can the effects of multiple institutions be compared because institutional provisions are not assigned randomly to treaties.

This chapter argues that accurately evaluating institutions requires paying more attention to why states design international institutions as they do and why, once institutions are negotiated, some states join and others do not. As Keohane and Martin note, “institutional theory needs to deal forthrightly with the endogeneity problem.”<sup>4</sup> We need to better distinguish between cases in which problem structure dictates institutional design (and therefore, the correlation of behavior with institutional rules is epiphenomenal) from those in which problem structure underdetermines institutional design (and, therefore, institutional design choices have the potential to exercise real influence over behavior). Making those distinctions convincingly requires identifying the influence of problem structure on institutional design, the behaviors that problem structure would dictate absent an institution, and the ways in which behaviors influenced by institutional rules are distinct and distinguishable from those influenced solely by the forces of problem structure.

The chapter first outlines how problem structure influences—but does not dictate—institutional design. It clarifies why (and the conditions under which) the “institution-independent” interests that states strive to codify in international institutions may diverge from the interests that subsequently

<sup>3</sup> Strange 1983.

<sup>4</sup> Keohane and Martin 2003, 73.

drive those states' behaviors. The chapter then outlines the ways in which problem structure and country characteristics also influence—but do not dictate—institutional membership. I then delineate methodological strategies for analyzing institutional effects that take design endogeneity and membership endogeneity into account. I provide a framework for analyzing institutional influence that recognizes that even institutions that reflect the interests and capabilities of the states involved nonetheless may lead those states to engage in different behaviors than they would otherwise. Using this framework, one can conduct research that would more convincingly distinguish whether actual behaviors reflect institutional influence or not.

### The Endogeneity of Institutional Design to Structure and Behavior

Both realist and institutionalist assumptions pose theoretical challenges to claims of institutional influence. I focus here on the influence of those international institutions that derive from treaties, conventions, protocols, and other formal agreements “negotiated among international actors, that prescribe, proscribe, and/or authorize behavior.”<sup>5</sup> Although international institutions can take many forms, the formal and public nature of international legal agreements facilitates analysis by removing analytic ambiguity about the commitments states have made to each other.

Realist claims of institutions as epiphenomena take two forms. The first critique is that international agreements reflect “wishful thinking” and a “false promise” on which they cannot deliver: states are no more likely to engage in the behaviors they commit to in international treaties than they would be in the absence of such commitments.<sup>6</sup> This critique predicts high levels of regime-inconsistent behavior whenever such behavior is inconvenient or costly. The second realist critique is that international agreements simply codify the existing constellation of power and interests among the states involved—states negotiate and accept only those rules that prescribe behaviors they would have engaged in anyway.<sup>7</sup> This view predicts high levels of regime-consistent behavior because such behavior is not costly. Thus, one interpretation of the high degree of Soviet and American compliance with arms control agreements during the Cold War is that those agreements simply put in writing previously planned weapons buildups rather than curtailing them. Likewise, the quotas established under the International Convention for the Regulation of Whaling throughout the 1960s and 1970s closely correspond to the catch in

<sup>5</sup> Koremenos, Lipson, and Snidal 2001, 762; Krasner 1983b.

<sup>6</sup> Mearsheimer 1994-95.

<sup>7</sup> See, for example, Werner and Yuen 2005.

the previous year, suggesting that catch was dictating quota more than quota was dictating catch. In the first critique, the behavioral dictates of structure and the behavioral dictates of institutional rules diverge, but behavior follows the former; in the second critique, structure dictates both institutional rules and behavior. In neither case, however, do institutions have any independent impact on behavior.

Institutionalist theory also poses obstacles to demonstrating institutional effects. Existing functionalist theories of the factors that determine why and when international institutions form and what dictates their terms imply "complete endogeneity" since they contend that "states create institutions when they expect them to be useful."<sup>8</sup> But if both the terms of international agreements and the behaviors those agreements regulate are influenced by problem structure, then research on institutional effects faces a troubling, omnipresent, and difficult-to-refute alternative hypothesis, namely, that any change in behavior after an international institution is created or amended is caused by a change in the underlying problem structure that leads states both to behave differently and to codify their intention to do so. Indeed, within this logic, the creation or modification of an international institution itself becomes evidence that the problem structure has changed; otherwise, why did states create or change the institution when they did? Thinking across institutions, this logic implies that variation in design simply reflects variation in underlying problem structure. The fact that one type of institution performs better than another stems not from one set of negotiators doing better at institutional design but from their facing an easier problem, that is, one more susceptible to institutional influence.

Those studying institutional influence and regime effectiveness have yet to respond adequately to these challenges.<sup>9</sup> To be sure, the literature on regime effects has grown considerably in the past three decades, not least through Keohane's extensive theoretical and empirical contributions.<sup>10</sup> This scholarship has often engaged the first realist critique, with many scholars quite carefully demonstrating that states behave consistently with institutional rules more frequently than structural factors would predict.<sup>11</sup> Scholars have also provided considerable evidence of Keohane's argument that the distribution of power differs across issue areas, with corresponding variation in the ability of states to influence outcomes (see Milner, introduction to this volume). Nonetheless, too little work has made

<sup>8</sup> Keohane and Martin 2003, 98 and 82. See also Keohane 1983, 1984.

<sup>9</sup> Keohane and Martin 2003, 98.

<sup>10</sup> See, for example, Keohane, Nye, and Hoffmann 1993; Keohane, Haas, and Levy 1993; Keohane and Levy 1996.

<sup>11</sup> See, for example, Brown, Weiss, and Jacobson 1998; Young 1999; Miles et al. 2002; Breitmeier, Young, and Zürn 2006.

evaluating the influence of structural and country-level factors on institutional design a central feature of efforts to evaluate how institutions influence behavior.

Scholars have begun addressing problems of membership endogeneity, that is, the difficulties arising because states that become members of an international institution differ in systematic ways *that influence their subsequent behavior* from those that do not become members. Thus, Simmons demonstrates institutional effects by contrasting behaviors of states that make international commitments and those that do not only after accounting for the factors that lead states to make those commitments.<sup>12</sup> Von Stein's research further highlights that institutional influence can be accurately assessed only after ensuring that members and nonmembers of international institutions do not differ in systematic but unobservable ways that lead to behavioral differences between the groups.<sup>13</sup>

The second realist critique, however, or the design endogeneity problem, has received far less attention to date, despite being evident since institutional theory was first developed. As Nye and Keohane note, precisely because "institutions often reflect the conditions of their origin, it is helpful to understand those origins."<sup>14</sup> Or consider Stein's claim that, because coordination problems pose inherently less risk of defection than do collaboration problems, the institutions states create to address the former will contain fewer, if any, monitoring and enforcement provisions than those they create to address the latter.<sup>15</sup> The arrangement of states' power and interests in coordination problems imply predictable post-agreement behaviors *regardless of the specifics of the agreement*; and these behaviors are predictably different from those in response to collaboration problems. Precisely because states recognize these predictable differences in postagreement behaviors, they incorporate different provisions in regimes addressing coordination and collaboration problems. The coordination/collaboration distinction provides merely one example of the larger point that differences in problem structure often imply quite different postagreement behaviors and that differences in institutional design reflect states' recognition of these differences.

Why has so little analytic effort addressed the design endogeneity problem? In part it is a by-product of the numerous factors that have led most scholars to investigate institutional effects via individual agreements. A focus on individual agreements results from several forces: a desire to engage realist claims of epiphenomenality with sufficient empirical richness to draw confident causal inferences; from a belief that "it is hard to find

<sup>12</sup> Simmons 2000.

<sup>13</sup> von Stein 2005; Simmons and Hopkins 2005.

<sup>14</sup> Nye and Keohane 1993, 111.

<sup>15</sup> Stein 1983.

enough comparable cases, across which institutional form varies but structure does not;<sup>16</sup> from the difficulty of identifying multiple institutions whose effects it makes sense to compare and for which comparable data is available; from a desire to control for variation in problem structure; and from methodological predilections. Even the best quantitative studies usually focus on a single agreement or even a single provision, comparing variation over time and across countries but not across institutions.<sup>17</sup> Whatever the reasons, design endogeneity is not particularly salient when evaluating a single agreement. Analyzing single cases, even cases that involve multiple observations, tends to control for problem structure, which facilitates isolation of institutional influence but precludes analysis of how variation in problem structure influences both institutions and outcomes. Fortna's work on peacekeeping missions, although not looking specifically at international agreements, engages problems that parallel design endogeneity by addressing how factors that influence the existence and type of peacekeeping missions deployed also influence the relative effectiveness of those missions.<sup>18</sup>

### Endogeneity and the Possibility of Institutional Influence

To the extent that institutional design is completely endogenous to structure, no room exists for institutional influence.<sup>19</sup> We therefore need a theory of institutions that recognizes and identifies the important ways in which structural factors influence both institutional design and institution-related behaviors but allows for the possibility that those factors do not dictate either. In a tension that reflects the larger theoretical debate regarding structure and agency, structure sometimes dictates institutional design but at other times allows negotiators flexibility to make meaningful choices about some aspects of institutional design.

I define *problem structure* as that array of factors which influences state behavior in such a way that at least some states view current outcomes as suboptimal and are, therefore, motivated to develop an international institution to improve them. Considered broadly, this definition includes realist factors such as the power and interests of states relevant to a problem, institutionalist factors such as preexisting institutions and rules, and constructivist factors such as ideas, norms, discourse, and identities. Taken together, these factors generate some set of "institution-independent" interests. I do not use the term *preinstitutional* interests since I want to

<sup>16</sup> Keohane and Martin 2003, 83.

<sup>17</sup> See, for example, Simmons 2000; von Stein 2005.

<sup>18</sup> Fortna 2004a, 2004b; but see also Werner and Yuen 2005.

<sup>19</sup> Keohane and Martin 2003.

capture the notion that these factors continue to influence interests even after an institution is established. In examining endogeneity, the relevant questions are how, why, and under what conditions might interests as “institutionally codified” diverge from institution-independent interests. Clearly, institutional influence is epiphenomenal unless such a divergence exists. Yet there are several reasons to believe that such divergence can occur.

First, to accept the proposition that problem structure influences institutional design does not require that one accept that it dictates it. Both theoretical reasons (e.g., uncertainty, bounded rationality, and unintended or unanticipatable institutional consequences) and empirical evidence (e.g., the time states take to negotiate agreements and the difficulty of predicting their terms until negotiations are complete) suggest that the interests that emerge from a given problem structure constrain but do not uniquely define an agreement’s features. Problem structure surely determines the “zone of possible agreement” but not what agreement—within that zone—states reach.<sup>20</sup> Put differently, most international institutions arise from non-zero-sum games whose solutions have multiple equilibria<sup>21</sup> in which the common desire to avoid suboptimal outcomes does not identify a unique point on the Pareto frontier.<sup>22</sup> Even some skeptics of institutional influence recognize that structure rarely uniquely dictates institutional form. Thus, the claim that international institutions that require “deep” cooperation, that is, that require significant behavioral change, will be ineffective unless they contain strong enforcement provisions implies that states can and do choose between strong and weak enforcement provisions in designing such institutions.<sup>23</sup> Structural factors and foresight surely incline states toward particular solutions to particular problems, but they can make both choices and mistakes in doing so. Constructivists further caution us that states’ interests neither preexist nor are fully exogenous to international negotiations—states may only define (and may redefine) their interests during negotiations.<sup>24</sup> At the extreme, if interests were perfectly known—and if they fully dictated institutional terms—then, presumably, negotiations would be short and their outcomes readily predicted, which they are not. Indeed, precisely because interests do not dictate institutional terms, states use negotiations as a means of defining and choosing among mutually acceptable alternatives. Because structure usually, though not always, gives states and their nego-

<sup>20</sup> Raiffa 1982.

<sup>21</sup> Keohane and Martin 2003.

<sup>22</sup> Krasner 1991.

<sup>23</sup> Downs, Rocke, and Barsoom 1996.

<sup>24</sup> Risse 2000.

tiators some freedom in choosing design features, “gaps will appear between problem and form.”<sup>25</sup>

Second, states are unlikely to be able to perfectly anticipate future situations and incorporate institutional design features that will foster their interests in those situations. If states can perfectly anticipate the future and can design institutions to achieve their preferred outcomes in those situations, then international institutions are epiphenomenal. But whenever the assumption of perfect anticipation seems unreasonable or when institutional design cannot be made flexible enough to address the expected range of contingencies, then institutions have the potential to exert influence, that is, to generate outcomes that would not occur otherwise. Indeed, Keohane identified this issue and its methodological implications over two decades ago in *After Hegemony*.<sup>26</sup> In that book and many others, Keohane’s methodological insight was that institutional effects are most visible not when regimes are first established but at subsequent points at which the original power and interests that led to, and are codified in, the institution’s design have shifted enough that the behaviors they would predict differ from those that the new constellation of power and interests would predict.<sup>27</sup>

Third, states cannot always accomplish their goals without institutions. Problem structures may predetermine particular outcomes in situations in which states cannot achieve their objectives by any means other than creating an institution. Thus, states may not be able to undertake an exchange that all sides prefer without creating a contract-like institution that helps reassure all sides that others will carry out their parts of the bargain.

Fourth, norms exist that constrain the institutional forms that states can use to codify their interests. Institutional provisions require relatively universal language, with assumptions that states have equal rights and obligations or that differentiation of rights and obligations must be based on general—not state-specific—criteria. In the absence of institutions, “the strong do what they can while the weak suffer what they must.”<sup>28</sup> By contrast, in negotiated institutions, the strong and the weak are presumed to deserve equal treatment. Processes for modifying collective obligations, opting out of such obligations, withdrawing from agreements, settling disputes, and so on, are standardized so that all parties receive equal treatment. Joint obligations and individual commitments are assumed to remain operative until explicitly revoked.<sup>29</sup> Although many

<sup>25</sup> Keohane and Martin 2003, 106.

<sup>26</sup> Keohane 1984.

<sup>27</sup> Keohane, Nye, and Hoffmann 1993.

<sup>28</sup> Thucydides 1982, book 5, chap. 89.

<sup>29</sup> Keohane and Martin 2003, 101.



agreements violate these metanorms, most follow them, suggesting that these metanorms influence which terms are incorporated—and which eschewed—in international agreements.

Finally, institutionally codified interests tend to diverge from institution-independent interests over time because the former tend to change more slowly than the latter. Even if negotiators could perfectly codify institution-independent interests in institutional rules, inertia prevents smooth and rapid institutional responses to changes in structural variables. Only “when institutions remain stable in the face of changes in structural variables, [do] we have an opportunity to observe the independent effect of institutions.”<sup>30</sup> This drift of institution-independent interests from institutionally codified interests also can arise in international agreements that establish institutional agents that have both autonomy from, and different interests than, the states they represent.<sup>31</sup>

These observations imply that those interested in institutional effects must distinguish the influence of “structure-through-institutions” from an institution’s independent influence. Those who contend that institutions are epiphenomenal see design endogeneity as complete, with problem structure perfectly embedded in institutional form and no room for any independent institutional influence. The argument just delineated leaves as an empirical question whether institutions exhibit complete or less-than-complete design endogeneity, and hence whether institutions can have any independent influence. To answer that question requires far more focus on the conditions under which, and extent to which, institutional designs vary, having controlled for problem structure. For some types of problems, we should expect problem structure to dictate—or very tightly constrain—institutional form, whereas for others, we might expect considerably more variation in institutional form in response to problems with essentially similar problem structures.

### Problem Structure and Endogeneity

Discussions of problem structure to date have focused on how states’ incentives and interests regarding a problem influence both the institutions they establish and their behavioral responses to them.<sup>32</sup> This section discusses specific examples of such influence and then delineates three additional parameters of variation in problem structure that have implications for both institutional design and behavioral responses.

<sup>30</sup> Keohane and Martin 2003, 101.

<sup>31</sup> Keohane and Martin 2003.

<sup>32</sup> Krasner 1983b; Koremenos, Lipson, and Snidal 2001; Hasenclever, Mayer, and Rittberger 1997.

### *The Incentive Structure of Problems*

Certainly, problem structures vary with regard to the interests relevant actors have both in seeing a problem resolved and in contributing to its resolution if agreement can be reached. Variation in these interests has implications for what institutions states negotiate, which states join those agreements, and how states that join them behave. As examples, consider the implications of the problem structures of upstream/downstream, coordination, and collaboration for institutional design, the subsequent behavior of relevant actors, and analysis of the influence of institutions. I use *relevant actors* to refer to those actors that the states concerned about a problem view as required for—or capable of contributing to—its resolution.

In upstream/downstream problems, the institution-independent interests of those causing the problem make them indifferent or even antipathetic to seeing the problem resolved. France's upstream position meant that their dumping of chloride into the Rhine was a problem for the Dutch, but not them.<sup>33</sup> Britain's position upwind meant that its emissions of acid precipitants were a problem for the Scandinavian countries but not for the British.<sup>34</sup> We might assume that the interests in such asymmetric externalities would preclude creation of cooperative institutional solutions. But, if downstream states (i.e., the "victims" concerned about the problem) would benefit enough from resolving the problem and have access to appropriate inducements, they can identify institutional terms that make upstream states prefer institutional membership to nonmembership. Institutions responding to such problems will reject sanctions in favor of carefully delineated exchanges in which concerned actors reward those causing the problem for changing their behavior.<sup>35</sup> Upstream states will only join agreements if they contain such rewards, a point illustrated by the refusal of China, India, and other developing countries to join the Montreal Protocol designed to protect the Earth's ozone layer until amendments established a developing-country-focused financial mechanism. It is quite clear that the "no-institution" counterfactual involves no behavior change: upstream states have no incentives to change their behaviors without remuneration, and downstream states have no incentives to offer remuneration unless there are institutional arrangements that protect them against renegeing by the upstream state. The design and degree of influence of institutions addressing upstream/downstream problems are likely to vary with the attractiveness, credibility, and contingency of the rewards being offered.

<sup>33</sup> Bernauer 1996; Bernauer and Moser 1996.

<sup>34</sup> Levy 1993.

<sup>35</sup> Mitchell and Keilbach 2001.

In coordination problems, most relevant actors have institution-independent interests in seeing a problem resolved and, once a solution is established, have little interest in defecting from that solution.<sup>36</sup> In coordination problems—as illustrated theoretically by Battle of the Sexes and empirically by ocean navigation and air traffic control—the advantages to relevant actors of *some* resolution to the problem lead them to conform to the rules of *almost any* resolution of the problem. Since relevant actors will either adopt the agreed-upon behaviors or clearly announce their intention to defect, institutional provisions may vary considerably in the detail needed to identify equilibrium behavior but will be quite similar in having few, if any, monitoring or enforcement provisions. In coordination problems, an institution's influence will be most visible when no institution-independent focal points exist: if states do not have the option to coordinate informally around some "natural" or hegemonically preferred focal point,<sup>37</sup> then institutions cannot simply codify behavioral preferences since those preferences are indeterminate prior to negotiations. In such cases, states want to coordinate their behavior but collective creation of an institution is necessary to identify how they should do so.

In collaboration problems, many and perhaps most relevant actors have institution-independent interests in seeing a problem resolved, but their interests in contributing to its resolution are institution-dependent since they depend on the probability of defection being detected and the costs of detected defection. Institutions designed to address such problems often entail reciprocal restraint and incorporate specific monitoring provisions linked to a system to induce credible and potent sanctioning for those who defect.<sup>38</sup> Yet problem structure allows institutional flexibility—Coasean remuneration or side-payments can be substituted for sanctions.<sup>39</sup> Indeed, as Keohane noted, states that are quite capable militarily may find it to be less costly to rely on "forms of power besides military force and threats."<sup>40</sup> Thus, in a 1911 agreement, the United States and Russia resolved a quintessential coordination problem involving open access to fur seal stocks by paying Canada and Japan to halt their fur seal harvests.<sup>41</sup> As in upstream/downstream cases, the resource transfers required careful monitoring provisions but provide empirical evidence that sanctions are not the only viable response to collaboration problems. Collaboration problems not only have alternative institutional solutions but may fail to be resolved because of the difficulty of collective action, particularly among

<sup>36</sup> Stein 1983.

<sup>37</sup> Lipson 1991.

<sup>38</sup> Downs, Rocke, and Barsoom 1996.

<sup>39</sup> Coase 1960.

<sup>40</sup> Milner, introduction to this volume.

<sup>41</sup> Slunaker 2005.

states that distrust each other. As with coordination problems, estimating the “no institution” counterfactual depends on determining whether the hegemon’s incentives were such that regime formation was a foregone conclusion and, if not, how the hegemon might have wielded its power to achieve the same objectives. When no hegemonic state exists, institutions that arise in response to collaboration problems among mutually distrustful states present strong counterfactuals: we can expect the problem to remain unresolved without an institution, and so any behavioral progress can more confidently be attributed to the regime. By contrast, highly interdependent states (e.g., European Union states) are more likely to create institutions to address their coordination problems but also more likely to make progress in mitigating the problem even without such an institution.

These three descriptions—although only a sample of problem structures—illustrate the larger point that variation in incentive structures produces predictable variation in institutional design, institutional membership, postagreement behavior, and no-institution counterfactuals. Problems, however, also exhibit variation in their capacity structure, informational structure, and normative structure, as discussed below.

### *The Capacity Structure of Problems*

The distribution of state capacities to contribute to problem resolution also varies across problems in ways that influence institutional form and effects. In some cases, positive interstate externalities may go unrealized if states lack the financial, technical, or administrative capacity to undertake behaviors that would benefit themselves and others. Consider the cases of drug interdiction, controlling AIDS and other diseases, and providing advanced military hardware to alliance partners.<sup>42</sup> In such cases, incentives still matter since states will not form international institutions if states that would benefit from other states addressing their own national problems do not see those potential benefits as greater than the costs they must incur to fix the other states’ incapacity problems. Yet resolving such problems requires institutions that address incapacities, not incentives, through transfers of knowledge, technology, or financial resources. Donor states will seek to minimize their costs and to ensure the transfers are used as intended. If they have only weak incentives to resolve the problem, they can dictate explicit terms of exchange, transfer less fungible resources, and limit the value of the transfers to the minimum needed to make the behavior possible. If donor states have strong incentives to resolve the problem, recipient states can extract more, and more

<sup>42</sup> See, for example, Darst 2001; Connolly and List 1996.

fungible, transfers. The strength of monitoring that states incorporate depends on whether the resources being transferred are relatively fungible or nonfungible, with agreements that transfer AIDS drugs likely to have weaker monitoring provisions than those that provide surveillance planes for drug interdiction. Recipient states are likely to be eager to join institutions addressing incapacity problems, whereas prospective donor states are likely to be reluctant to do so. Problems for which incapacity is central provide clear no-institution counterfactuals that differ markedly from postagreement outcomes if the institution is influential: states that are incapable of an activity and of developing the capability indigenously but begin engaging in that activity in ways that institutional resource transfers made possible provide strong evidence of institutional influence.

### *The Informational Structure of Problems*

Institutional form and effects are also influenced by a problem's informational structure. Behaviors that give rise to interstate problems vary considerably in how visible they are and how well they can be kept secret. Thus, arms races and tariff wars are usually considered to be prisoner's dilemma problems, but vary considerably in their institution-independent transparency: states have strong incentives and often the ability to keep military activities secret, but they cannot collect tariffs without revealing them to actors that have strong incentives to report rates that exceed institutionally sanctioned levels.

Institution-independent transparency varies within as well as across issue areas: compare tariffs and quotas to nontariff barriers; large-scale troop deployments to nuclear weapons development; or denying equal suffrage to torturing prisoners. To be influential, institutions addressing the latter of each of these pairs must craft far more careful monitoring provisions. Indeed, inherently nontransparent behaviors tend to prompt institutional innovation to enhance transparency, as evident in the surprisingly cooperative procedures for on-site inspection developed in U.S.-Soviet nuclear arms agreements and in states facilitating monitoring by nongovernmental actors in areas of human rights that governments are less capable and less inclined to monitor.<sup>43</sup> When states develop careful and cooperative monitoring provisions in response to inherently nontransparent activities, such as International Atomic Energy Agency inspections, the no-institution counterfactual claim that those activities otherwise would not have come to light and, therefore, would not have changed, gains plausibility. By contrast, claims of influence for institutions addressing inherently transparent activities are more challenging

<sup>43</sup> Keck and Sikkink 1998.

and must demonstrate that observed behavioral responses were not a response to the institution-independent existence of that information.

### *The Normative Structure of Problems*

Finally, problems vary in normative structure. Institutional design and postagreement behaviors vary considerably depending on when in a normative life-cycle states succeed in creating an international institution.<sup>44</sup> At one end of the spectrum, states may negotiate international agreements only after both norms and behaviors have stabilized, as in the codification of customary international law. At the other end, nonstate activist networks may successfully press states to negotiate agreements designed to promote a normative consensus where none yet exists, as in signature of the land mines convention at a time at which most nations still viewed them as legitimate weapons.<sup>45</sup> In the spectrum's middle are many cases in which states that have accepted certain norms create international institutions to induce others to do so.

Normative variation also exists along a spectrum from issues in which interests are at stake and norms play little role to those in which norms are central. Thus, most arms control, trade, and natural resource negotiations involve interest-based dialogues based in a logic of consequences. By contrast, most human rights negotiations involve norm-based dialogues based in a logic of appropriateness.<sup>46</sup> We can expect institutions designed to reduce behaviors already viewed as morally "bad"—such as torture or genocide—to specifically define what constitutes a violation and to rely on sanctions for violations. By contrast, we can expect institutions designed to foster behaviors viewed, by some, as morally "good" but insufficiently frequent—such as equal treatment of women and minorities or ensuring citizens' economic and social well-being—to be broader or more exhaustive in their definitions of commitments. To promote internalization of norms, these latter institutions will tend to foster extended normative dialogue, persuasion, argument, and the praise of those doing well rather than instrumental mechanisms such as sanctions or side-payments, since the latter undermine efforts to induce the internalization of norms.<sup>47</sup>

Normative variation also has implications for estimating counterfactuals: the later in a normative life-cycle that an international institution is established, the more difficult it becomes to plausibly claim that behavioral conformance with institutional rules is due to the institution. Equally

<sup>44</sup> Finnemore and Sikkink 1998.

<sup>45</sup> Price 1998.

<sup>46</sup> March and Olsen 1998.

<sup>47</sup> Risse 2000.

important, the longer after the creation of a norm-based institution that one attempts to assess institutional influence, the more difficult it becomes to disentangle whether normative strengthening is caused by institution-independent normative maturation or by the institution's contribution to that maturation. Thus, in most human rights cases, the appropriate no-institution counterfactual is probably to assume that the norm would have gained some strength even absent any institution; evidence for institutional influence must come from a rate of norm internalization and acceptance that is higher than what would have been expected otherwise.

### *Other Considerations*

Problems also vary with respect to whether they involve acts of omission, acts of commission, or accidents; whether states and civil society are tolerant or intolerant of violations; whether a consensus does or does not exist about the problem and its resolution; and whether resolution of the problem requires dramatic and costly behavioral changes or all-but-costless changes. Variation in these and other parameters of problem structure may well influence both institutional terms and state behaviors in ways that pose analytic risks of endogeneity and hence require attention when evaluating institutional influence.

### *Multifinality and Equifinality*

The foregoing discussion of a subset of different problem structures illustrates that the relationship of problem structure to institutional design can involve both multifinality and equifinality. In multifinality, negotiators have sufficient freedom to incorporate different institutional features in response to a given problem structure; in equifinality, negotiators addressing problems with different structures may end up selecting similar institutional features. Both aspects of the relationship provide analytic opportunities: situations of multifinality allow evaluation of the relative effectiveness of different institutional design elements, with the investigator controlling for problem structure; situations of equifinality allow comparison of the performance of a given institutional design element in different problem contexts.

### Methodological Implications

Recognizing the endogeneity of institutional design to problem structure generates several methodological insights. Those interested in identifying institutional effects must avoid assuming that problem structure and in-

stitutional design are perfectly correlated or, conversely, assuming that they are totally independent. If design endogeneity is likely but neither necessary nor perfect, then it is an empirical question whether a given institution's design is endogenous to problem structure. Those interested in institutional effects cannot take institutional design features as given but must examine the extent to which those features were independent of problem structure. Keohane's dictum to "think like a social scientist"<sup>48</sup> suggests researchers must carefully identify the values of various parameters of variation in problem structure independent of evidence regarding institutional change (not by taking institutional change as evidence of problem structure change) and then examine the extent to which variation in problem structure covaries with institutional design.

Making a convincing claim of institutional influence requires explicitly demonstrating that institutional design was not completely dependent on problem structure, whether by demonstrating that problem structure was indeterminate with respect to institutional design or by demonstrating the presence of institutional variation in the absence of problem structure variation. Researchers also must avoid assuming that—and instead carefully evaluate whether—an institution's members and nonmembers are otherwise alike. As noted, some scholars have begun to address membership endogeneity.<sup>49</sup> As with design endogeneity, the extent of membership endogeneity is an empirical question that cannot be identified by assumption. When comparing the behavior of members to nonmembers or of states before and after becoming members, scholars must try to identify what factors make states likely to join particular agreements at the time that they do independent of evidence regarding when they joined them.

Analysts should also explicitly consider what alternatives to the observed institution, if any, were possible given the problem structure. This requires assessing theoretical arguments and empirical evidence to determine how much leeway, if any, negotiators had in institutional design. How "inevitable" was it that states would create an institution and that the institution would take the form it did? Knowing whether states might have forgone the institution and what alternative institutions they might realistically have created is crucial to generating plausible counterfactuals that are the foundation for claims of institutional influence.

Addressing both design endogeneity and membership endogeneity requires methods that distinguish an institution's independent influence on behavior from both problem structure's direct influences on behavior and its indirect influences through its influence on institutional design. One approach involves comparing cases for which the analyst has verified,

<sup>48</sup> King, Keohane, and Verba 1994, 32.

<sup>49</sup> Simmons 2000; Simmons and Hopkins 2005; von Stein 2005.



rather than assumed, that problem structure was held constant. Whether looking for the effects of changes in a single institution's design over time (e.g., adoption of protocols or amendments) or the effects of institutional variation across institutions, the possibility of endogeneity demands that the analyst explicitly evaluate the values of problem structure parameters and demonstrate that they do not vary. A second approach, which addresses how well a given institutional design does in varying circumstances, involves comparing cases that exhibit a change in problem structure but institutional constancy, the methodological underpinning of Keohane's argument in *After Hegemony*.<sup>50</sup> Here, too, the analyst must clearly demonstrate that problem structure changed in ways that would be expected to produce an institutional change but did not.

With respect to design endogeneity, analysts should start by assuming that some part of any correlation between behavior and institutional design arises because both are driven by problem structure. With respect to membership endogeneity, analysts should start by assuming that member states are more likely to adopt institution-consistent behaviors than non-member states even in the absence of an institution, not because of institutional influences but because states that become members are more concerned about the problem, face lower adjustment costs, or both.<sup>51</sup> Analysts should begin with these assumptions but then look explicitly for variation in the problem structures being addressed or in the characteristics of states that joined an institution and those that did not. Institutional influence can be shown either by identifying behaviors that differ systematically and in predicted ways between the institutional and noninstitutional setting or by finding evidence that the processes and causal mechanisms that led to certain behaviors more closely match the predictions of an institutional argument than those of a problem structure argument.

Quantitative analysis of institutional effects opens up additional possibilities, so long as the foregoing arguments about endogeneity are taken into account.<sup>52</sup> Standard regression techniques separate institutional influences from structural influences but do not usually separately identify structure-through-institution influences. Design and membership endogeneities require explicit modeling. A particularly appropriate approach involves two-stage modeling in which the influence of structure on institutional design is determined in one regression and the results of that regression are used as inputs to a regression that assesses the influence of institutional design and structure on behavior, having already accounted for the influence of structure on institutional design. These and related

<sup>50</sup> Keohane 1984.

<sup>51</sup> Sprinz and Vaahoranta 1994.

<sup>52</sup> Miles et al. 2002; Helm and Sprinz 2000; Hovi, Sprinz, and Underdal 2003a, 2003b; Young 2003; Mitchell 2002.

strategies offer opportunities to evaluate the extent to which institutional design reflects or diverges from the dictates of problem structure; by accounting for that variation in analyzing behavioral responses, they can distinguish the influence of problem structure *through* institutions from the independent influence *of* institutions.

## Conclusion

Endogeneity of institutional design raises an analytic obstacle that has received insufficient scholarly attention. The same states whose (presumably self-interested) behavior is leading to suboptimal outcomes negotiate agreements (that presumably reflect their self-interests) to regulate those behaviors. If the international institutions states devise to regulate behavior and the behaviors being regulated are simply different expressions of the same constellation of power and interests, how can those institutions be said to have an independent influence on state behavior? The proper question is, as Keohane has argued in many settings, not either/or but "Under what conditions?" Institutions do reflect the power and interests of the states that create them, but they also can wield independent influence over those states' behaviors. Our task is to determine, convincingly, the conditions under which the latter happens given the methodological obstacles posed by the former.

The processes and constraints of creating international institutions produce institutional designs that can diverge, at least at times and to some extent, from what problem structure alone would dictate. That said, some problem structures may generate perfect design endogeneity, and, in those cases, any correlation of behavior with institutional dictates is epiphenomenal. But when structure underdetermines institutional design (or even whether states will create an institution), the collective institutional choices that states make may produce outcomes quite different from those predicted by problem structure alone. Distinguishing these two types of cases requires considerably more attention than recent institutional scholarship has given to three analytic tasks: (1) identifying how problem structure influences institutional design, (2) identifying what behaviors structural forces would dictate absent an institution, and (3) identifying how the behavioral dictates of institutions diverge from those of structural forces. In short, those interested in institutional effects cannot ignore the factors that influence institutional formation. It is only by recognizing the theoretical and methodological implications of endogeneity that scholars can take advantage of the many insights that Robert Keohane has contributed to our understanding of institutional influence over the past quarter-century.