Part Two: The Effectiveness of International Environmental Regimes

# Problem Structure, Institutional Design, and the Relative Effectiveness of International Environmental Agreements

*Ronald B. Mitchell**

## Introduction

Assessing the relative effectiveness of international environmental agreements accurately requires careful attention to the structures of the problems they address. Although even studies of single agreements would benefit by evaluating the influence of problem structure, problem structure poses particularly large analytic challenges for scholars interested in the relative effectiveness of environmental treaties. To determine which aspects of institutional design account for the better performance of some environmental agreements requires methodological choices that address the influence of problem structure on both design and the behaviors agreements target.

Scholars are increasingly undertaking comparative analyses of environmental agreements, regimes, and institutions, to extend and improve our understanding of how much different institutional design elements contribute to an agreement's effectiveness. The need to account for problem structure when analyzing institutional influence arises from four factors. First, problem structure is

* Comments from Steven Rothman, Erin Graham, and three anonymous reviewers as well as research assistance from Claire Tonry challenged the author to make this argument clearer and stronger than it would have been otherwise. The material in this article 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 2006. 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.

*Global Environmental Politics* 6:3, August 2006

a major alternative explanation to institutional influence. Second, institutional influence may depend on interactions between institutional design elements and problem structure variables that serve as conditioning factors. Third, institutional design is not independent of but is endogenous to problem structure, i.e., problem structure influences both the type of institutions states create and how likely states are to respond to those institutions. Fourth, variation in problem structure influences how states define success, which should be reflected in the metrics of institutional effectiveness that we adopt. Recent scholarship has begun to recognize the importance of the first of these two problems but has been less attentive to the latter two. Remedying existing shortcomings in this regard requires identifying those aspects of variation in problem structure that influence institutional design, controlling for such variation in selecting cases for comparison by using empirical distinctions to capture theoretically-informed variation, looking for remaining variation in these variables, evaluating how both the problem structure variables successfully controlled for and those that still vary influence both institutional design and behavioral outcomes, and using problem structure variables to identify the most appropriate metric of institutional effectiveness. These strategies could be used to select and compare various subsets of international environmental agreements but are illustrated here through identifying methods for conducting an analysis of high seas tuna fisheries. This article builds on elements central to Arild Underdal's impressive research legacy. Underdal has led efforts to assess environmental regime effectiveness systematically.[1] He has shown that comparing agreement effects, determining the effects of particular design features, and assessing their generalizability can be fostered by supplementing qualitative studies of single agreements with quantitative assessments of multiple agreements.[2] His recent work has highlighted the centrality of problem structure to analysis of institutional effectiveness.[3] More than most, he has thought systematically and quantitatively about the effects of regimes, developed sophisticated methods for analyzing them, and conducted particularly nuanced and careful empirical work.[4]

## The value of comparing agreements

Analysis of the effectiveness of international environmental agreements and institutions—for current purposes, I use the terms interchangeably—has reached a point of theoretical development and empirical knowledge that we can move beyond questions of whether particular agreements had influence to which agreements had more influence and why. One important focus of Underdal's re-

1. See Underdal 1992; Underdal 1998; Miles et al. 2002; and Underdal and Young 2004.
2. Underdal 2004, 43.
3. Underdal 2002b.
4. See, Underdal 2004; Underdal 2002a; Underdal 1980; but, see also, Sprinz and Helm 1999; Mitchell 2002; Hovi, Sprinz, and Underdal 2003; and Young 2003.

search has been the relative effectiveness of different institutional arrangements, i.e., "so, which worked better?"[5] Convincing answers to this question cannot come from the qualitative case studies or game theoretic analyses that have dominated the field to date.[6] Underdal's pathbreaking cross-case analysis in the concluding chapter of *Environmental Regime Effectiveness* complements a growing body of comparative work.[7] Most of this research compares the effectiveness of agreements that address environmental problems with quite different problem structures, and rely on often-different experts to assess each institution's influence, how much influence each had, and what aspects of each best account for its influence. This article affirms the value of comparative analysis of effectiveness but proposes that efforts to determine "which worked better" could be improved by taking problem structure more seriously than has been done to date.

Analyzing multiple agreements offers three benefits over analyzing single agreements. Comparisons allow us to move more confidently from claims that an agreement was influential to claims regarding which variables, of the many proposed by scholars, explain such influence. In a single case, it is often difficult to determine which of several institutional features or other factors, all of which were present, actually explains observed outcomes. Case selection, process tracing, examination of within-case variation, and other strategies can ease these problems. Comparing multiple cases, however, increases our ability to identify confidently whether the institution or non-institutional factors caused observed behavioral variation and, if the institution was responsible, which institutional features were "active ingredients" in the agreement's success and which were superfluous.

Analyzing multiple agreements also increases opportunities to assess how institutional influence depends on non-institutional conditions. For any given agreement, it is often difficult to exclude the possibility that the agreement's effectiveness was contingent on a set of "benign" facilitating factors.[8] Designs that work well under certain conditions may work poorly or be counterproductive under other conditions.[9] Examining several agreements allows us to determine whether the influence of particular institutional features are unique to particular contexts, are systematic across contexts, or are fostered or inhibited by particular aspects of the context. Relatedly, it can clarify the average effect of particular features across a range of contexts.

Third, comparing multiple agreements fosters assessment of the influence of a particular design feature both to other features and to factors outside the control of negotiators. Thus, comparative studies offer opportunities to deter-

5. See, in particular, Miles et al. 2002.
6. Underdal 2004, 43.
7. See, for example, Brown Weiss and Jacobson 1998; Victor, Raustiala, and Skolnikoff 1998; Young 1999a; and the International Regimes Database project describe in Young and Zürn, this issue.
8. Underdal 2004.
9. Underdal 2002a, 449.

mine which of several institutional options (such as sanctions, capacity enhancements, or performance-based rewards) work best, how contingent their influence is on other institutional features (such as monitoring), and how much influence they wield relative to non-institutional features (such as changes in prices or technologies).

## The Need to Account for Problem Structure

The ability to draw compelling conclusions from comparisons of multiple agreements depends, however, on making problem structure central to the analysis for three reasons. First, "problem structure" captures a wide array of alternative explanations to institutional influence. As Underdal has recently noted, institutional performance depends on both problem structure and institutional design.[10] International institutions may perform well because they are well designed or because they address easy problems. Even if we assume, initially, that institutional influences are truly independent of problem structure, institutional influences cannot be accurately identified without analytically removing the influence of problem structure. Both policy-makers and scholars may want to know whether particular problems are conducive or resistant to institutional influence, not least to estimate the relative resources needed to address them successfully,[11] but may also want to know the relative effectiveness of alternative institutional designs at remedying a given problem. In either case, problem structure requires careful analytic attention.

Second, there may be interactions between institutional design and problem structure variables, with a particular design feature's influence depending on the problem structure being addressed. Thus, agreements that lack sanctions may resolve problems requiring shallow cooperation effectively but not those requiring deep cooperation.[12] Collaborative environmental monitoring may contribute significantly to agreements among industrialized states but far less to those among developing states. To identify the expected effects of institutional features under different conditions and to judge the generalizability of analytic findings requires knowing either the values of problem structure variables or that they have been controlled for.

Third, and central to the argument here, strong theoretical reasons exist to believe that institutional design is endogenous to, rather than independent of, the relationship between problem structure and behavior. Endogeneity involves a relationship in which, of two independent variables that are claimed to influence a dependent variable, one influences the other.[13] International relations scholars have increasingly recognized that the same problem structure

10. Underdal 2002b, 3, 13.
11. Young 1999b, 51.
12. Downs, Rocke, and Barsoom 1996.
13. On issues of endogeneity, see, Keohane and Martin 2003; von Stein 2005; and Simmons and Hopkins 2005.

variables identified as the sources of institutional design (i.e., regime design variables) are also identified as the sources of the behaviors those institutions target (i.e., regime effectiveness variables).[14] Endogeneity implies that we should assume (until proven otherwise) that variation in institutional design is not independent of but is actually due to variation in underlying problem structure. An initial assumption that institutional design and problem structure are independent, noted above, corresponds to a "no endogeneity" assumption. At the opposite, "perfect endogeneity," extreme, if an institution's design is dictated (rather than merely influenced) by problem structure, then the institution cannot be said to have any independent influence on behavioral outcomes.[15] Under conditions of perfect endogeneity, international institutions are epiphenomenal[16] with problem structure explaining all variation in outcomes, either by directly influencing those outcomes or by indirectly dictating institutional design. The analytically bothersome aspect of endogeneity is that controlling for problem structure no longer constitutes an adequate method for rejecting the argument that outcomes were influenced by it. If we cannot assume institutional design and problem structure are independent, it seems reasonable to initiate analysis with an empirically agnostic belief that, in any given comparison of cases, we must assess whether problem structure a) dictated that states would reach agreement on a problem and dictated the major design features of that agreement, b) dictated that states would establish *some* institution but left negotiators leeway with respect to institutional design, or c) simply made it likely that states would establish an institution but left open the possibility that states might fail in that effort.

Consider how the number of actors engaged in a targeted activity affects assessment of the relative effectiveness of the Montreal Protocol on Substances that Deplete the Ozone Layer and the Convention on International Trade in Endangered Species (CITES).[17] "The limited number of facilities that produce ozone-depleting substances and the millions of individuals who could engage in illicit trade in endangered species helps to explain why CITES was much more difficult to enforce than the Montreal Protocol."[18] But the recognition of this difference in the underlying problems also influenced the roles states allowed NGOs and multinational corporations to play in monitoring and the types of enforcement provisions states adopted.[19] Here, the same problem structure variable (number of actors) explains variation in how likely states are to change their behavior in response to these agreements and the mechanisms by which these agreements target behavior. This makes it inherently challenging to

14. Koremenos, Lipson, and Snidal 2004.
15. Keohane and Martin 2003.
16. Strange 1983; and Mearsheimer 1995.
17. On number of actors involved in an activity, see Jacobson and Brown Weiss 1998a; and Koremenos, Lipson, and Snidal 2004.
18. Jacobson and Brown Weiss 1998a, 521.
19. Jacobson and Brown Weiss 1998a, 526–528.

distinguish whether variation in the behavioral influence of the agreements reflects truly independent differences in institutional design or differences in problem structure that happened to express themselves in differences in institutional design.

Consider the research design implications if endogeneity were perfect, i.e., if one could perfectly predict the design of any institution if all aspects of problem structure were known. Selecting cases to ensure variation in institutional design would necessarily mean one had simultaneously ensured variation in problem structure, and could therefore not determine whether the problem structure or the institution were the source of any observed variations in behavior. Similarly, selecting cases to control for problem structure would eliminate variation in institutional design. Accepting that endogeneity is possible does not imply accepting that it is perfect, however.[20] Indeed, perfect endogeneity implies that states and their negotiators exercise no agency in designing international institution. A more reasonable position would seem to be that states facing a set of problems with similar structures have choices among real, but constrained, alternatives. We can expect that cases selected to control for problem structures will exhibit sufficient variation in institutional design to warrant study. Indeed, the degree of institutional design variation across cases in a study that carefully and successfully selected cases to control for problem structure would provide compelling evidence to refute the "perfect endogeneity" claims of some realists.[21] That said, those institutional features that were common to those agreements would need to be presumed as dependent on the problem structure, demonstrating both that alternatives to those elements of institutional design were probably not truly viable and suggesting caution in generalizing to agreements facing different problem structures.

Finally, problem structure has important implications for how we define and evaluate institutional effectiveness. The nature of problems and the ways states perceive and define them influence what behavioral and environmental goals they establish in international institutions. Whether states set deep or shallow cooperative goals depends on both the collective motivation to resolve the problem and the availability of alternatives. States may create institutions that ban environmentally harmful behaviors when reasonable-cost alternatives exist but that only restrain growth in those behaviors otherwise. The extent of threats to different species and the nature of human concern have produced agreements that ban harvest of some species, severely restrict harvest of others, and only manage harvest of yet others. Resolving some environmental problems only requires that states regulate the level of particular activities; resolving others requires that states regulate the ways states engage in particular activities. Such differences imply quite different patterns of what we expect to see if the

20. Keohane and Martin 2003.
21. Strange 1983; and Mearsheimer 1995.

agreement is effective and quite different counterfactuals to use in estimating that effectiveness.

## Problems with Past Efforts

Considerable room exists to improve how comparative studies account for problem structure. Oran Young, Edith Brown Weiss and Harold Jacobson, and Edward Miles and Arild Underdal have published three superb comparative studies of regime effectiveness.[22] All three edited volumes provide high-quality comparative analyses of different regimes, combining detailed case studies of variation within regimes and overarching assessments of variation across regimes. All three, in varying degrees, include problem structure in their analyses. Nonetheless, these volumes illustrate shortcomings in incorporating problem structure in effectiveness research.

One shortcoming is simply the lack of consensus, reflecting the broader international relations literature, regarding the importance and proper way to describe problem structure. Although Young identifies important elements of problem structure variation in other work,[23] Young's comparative study identifies problem structure as important to evaluating institutional effectiveness but provides little sense of the ways in which it varies.[24] Brown Weiss and Jacobson do not use "problem structure" terminology but link institutional effectiveness to "four characteristics of the activity involved" and twelve "factors involving the country."[25] Miles, Underdal, and colleagues carefully identify the incentives of the underlying game, asymmetries, and cleavages as determinants of problem malignity, and couple that with uncertainty and "problem solving capacity" as the major aspects of problem structure. International relations theory provides a plethora of distinctions in problem structure including coordination vs. collaboration vs. suasion games, problems requiring deep or shallow cooperation, and conflicts over absolutely assessed goods, relatively assessed goods, ends, or means.[26] A major recent project identified the central elements of problem structure as distribution problems, enforcement problems, the number of actors and asymmetries among them, and uncertainty.[27] Thus, the current state of research makes it unclear how central problem structure should be to a study, what elements are most important to include, and what terms to use in doing so.

These volumes also illustrate a lack of clarity about the appropriate level of resolution in addressing problem structure. Young's choice to discuss problem

22. Brown Weiss and Jacobson 1998; Young 1999a; and Miles et al. 2002.
23. Young 1999b, chapter 3.
24. Young 1999c, 272–274; and Young 1999b, chapter 3.
25. Jacobson and Brown Weiss 1998a, 536.
26. Stein 1983; Miles et al. 2002; Martin 1992; and Rittberger and Zürn 1990.
27. Koremenos, Lipson, and Snidal 2001.

structure only in his volume's conclusion suggests that we can draw confident conclusions about regime effectiveness with only high-level and largely inductive distinctions among problem structures.[28] Underdal's framework relies extensively on the aggregate malign/benign distinction, the simplicity of which Underdal recognizes but defends as appropriate on empirical grounds.[29] In the present context, the significant problem[30] with Underdal's aggregate notion of malignity is that the aggregation facilitates identifying the difficulty of resolving problems and the factors necessary to promote success but loses the advantages of keeping problem structure variables disaggregated so they can be used to make specific predictions about institutional design. In contrast with both volumes, Brown Weiss and Jacobson identify at least sixteen different aspects of problem structure variation, many of which could be used to predict institutional design, although they do not fully develop such predictions.[31]

None of these volumes effectively address variation in problem structure in line with the issues delineated in the previous section. Young and colleagues focus on behavioral pathways and causal mechanisms rather than problem structure in their analysis. Brown Weiss and Jacobson make problem structure variables important aspects of their analytic framework, but focus on the independent—while ignoring the interactive and endogenous—aspects of those variables. Underdal and his colleagues make a commendable effort to examine effectiveness after controlling for problem structure.[32] Notably, none of these studies consciously control for problem structure. Indeed, Brown Weiss and Jacobson deliberately selected five treaties that included both pollution and natural resource protection cases,[33] and Miles, Underdal and their colleagues selected cases to ensure variation in, inter alia, problem structure and distribution of power.[34] All three studies select cases to be representative of environmental regimes more generally, which has its merits. But, selecting cases in which both problem structure and institutional design vary inhibits efforts to assess which best explains variation in behavioral outcomes.

These analytic problems reflect, in part, two related obstacles to current research on institutional effectiveness. First, empirical operationalizations for most theories regarding problem structure variables are rudimentary, at best. Even distinctions that are theoretically clear in two player game matrices, such as coordination/collaboration distinction, prove difficult to apply to the multiplayer, multi-motive contexts that characterize many environmental problems. Second, the absence of a systematic listing of environmental agreements and the structure of the problems they face (which, in part, reflects the just-discussed

28. Young 1999c, 272.
29. Underdal 2002b, 21–22.
30. Young 1999b, 58ff.
31. Jacobson and Brown Weiss 1998a.
32. Underdal 2002a.
33. Jacobson and Brown Weiss 1998b, 14.
34. Underdal 2002b, 37.

lack of agreement on how to categorize problem structure) has led many scholars to assume that few opportunities exist to compare essentially similar environmental problems.

## Improving our Approach to Problem Structure

Foregoing sections have identified the importance of incorporating problem structure into analyses of institutional effectiveness and the shortcomings of previous efforts to do so. This section delineates several analytic steps that would improve our efforts to address the influence of problem structure so that we can draw more compelling and nuanced conclusions about institutional effectiveness.

### *Describing Variation in Problem Structure: Incentives, Capacities, Information, and Norms*

Problems related to selecting the terms and level of resolution at which to describe problem structure suggests that properly accounting for problem structure requires identifying its distinct aspects at a level at which existing theory allows us to deduce implications for both institutional design and targeted behaviors. Evaluating the various taxonomies of variation in problem structure with the goal of identifying aspects that influence institutional design and targeted behaviors suggests four broad categories of variables. Distinctions based on incentives, particularly as captured in different games, have dominated much discussion of problem structure. Distinctions among coordination, collaboration, and upstream/downstream problems, though not exhaustive, seem particularly pertinent to environmental regimes and provide strong predictions about institutional design and targeted behaviors: states will craft more careful monitoring provisions for collaboration and upstream/downstream problems than for coordination problems, will tend to use sanctions for collaboration problems and rewards for upstream/downstream problems, and will be progressively less likely to meet agreed-upon obligations as one moves from coordination to collaboration to upstream/downstream problems.[35]

Capacities have been central to many analyses of environmental regime effectiveness. States often engage in environmentally harmful behaviors—and fail to comply with international efforts to restrain those behaviors—because of incapacities rather than intention.[36] Thus, many developing countries fail to protect the health of their populations adequately because of financial, administrative, and technical incapacities. Variation in capacities also influences insti-

35. Stein 1983; Martin 1992; Koremenos, Lipson, and Snidal 2001; and Mitchell and Keilbach 2001.
36. Chayes and Chayes 1995; Jacobson and Brown Weiss 1998a; and Keohane, Haas, and Levy 1993.

tutional design: states are more likely to rely on financial assistance and capacity enhancements rather than sanctions when they consider a problem to involve incapacity rather than incentive problems.[37]

The informational environment is also central to problem structure. Uncertainty about the state of the world and about the behaviors and preferences of other actors contribute to many international problems.[38] Greater scientific uncertainty about an environmental problem leads states to be more reluctant to alter their behaviors but prompts them to make scientific components more central to any institutions they may establish. Likewise, the less transparent one state's behaviors are to others, the more likely it is that those so inclined will engage in those behaviors and the more likely it is that states will carefully craft institutional monitoring provisions.[39]

Finally, the normative setting constitutes an additional aspect of problem structure. States may develop institutions late in a normative lifecycle, where those behaviors already receive considerable support from non-institutional norms. On the other hand, institutions can develop early on in a normative lifecycle, facing the need to develop a weak or create a non-existent norm and/or to try to discourage behaviors that are supported by countervailing norms.[40] Where norms regarding an environmentally-harmful behavior are strong (and countervailing norms are absent), states are likely to accept deeper, more demanding requirements while also being less likely to engage in those behaviors in the first place.

These four categories do not exhaust important distinctions in problem structure. Other important distinctions include the number of actors involved, the asymmetry in power and incentives among actors, and the positions taken by nongovernmental organizations, multinational corporations, and civil society with respect to the problem. The point here has been to identify elements of problem structure whose implications for both institutional design and targeted behaviors highlight the endogeneity problem that has so often been ignored.

### *Selecting Cases to Limit Variation in Problem Structure*

Once important variation in problem structure has been identified, research interested in evaluating institutional effects should select cases for comparison that limit variation in problem structure. Many scholars assume—based on such cases as stratospheric ozone loss, climate change, or hazardous waste—that there simply are not agreements that are sufficiently similar to warrant comparison. But in several arenas states have signed multiple agreements that address

37. Chayes and Chayes 1995; Downs, Rocke, and Barsoom 1996; and Keohane, Haas, and Levy 1993.
38. Koremenos, Lipson, and Snidal 2001; and Underdal 2002b.
39. Jacobson and Brown Weiss 1998b, 6.
40. Finnemore 1996.

relatively similar problems—consider marine fisheries, international lakes and rivers, regional seas, and port state control of marine pollution.[41] Using such empirical distinctions in problems is clearly second-best to using theoretical distinctions such as those noted above. But, no systematic database of agreements' problem structures (based on the foregoing or other taxonomies) exists from which one could select cases that share problem structures. Indeed, considering how to convincingly measure many aspects of problem structure suggests that there are significant obstacles to creating such a database. In the absence of such a database, well-chosen empirical distinctions provide useful proxies for theoretical distinctions.

Two initial steps to controlling problem structure involve excluding cases that address empirical problems so starkly different that they seem unlikely to share problem structures and then evaluating remaining cases to further eliminate variation in important aspects of problem structure in the cases finally selected. Thus, we might initially distinguish between pollution problems and natural resource problems,[42] dividing the latter further into natural resource management and natural resource conservation problems. Natural resource management problems include fisheries as well as management of fur seals, polar bears, whales, other endangered species, tropical timber, and wetlands. The large number of international fisheries agreements—states have negotiated over 50 multilateral agreements directly regulating fisheries since 1950—allows restriction of cases to high seas fisheries. Although this approach selects cases in or out based on theoretically-thin empirical criteria, these criteria serve as useful proxies for more analytic distinctions. Thus, overfishing reflects, almost quintessentially, collaboration (tragedy of the commons) incentives that are thought to make cooperation difficult.[43] High seas fisheries, on which Underdal has done significant analyses,[44] also present "hard cases" in another important respect: almost all commercially viable fish stocks are at least 90% below pre-exploitation levels.[45] Comparing hard cases reduces our uncertainty in estimating counterfactuals, reduces the number of rival explanations for any observed changes in behavior, and makes generalization to "easier" cases more plausible.[46] Excluding bilateral agreements addresses the "number of actors" aspect of problem structure.[47] Focusing on high seas cases focuses research on cases in which states engaged in the targeted behavior clearly have the capacity to stop doing so and those desiring to limit the activity have neither the legal nor practi-

41. Mitchell 2003.
42. Jacobson and Brown Weiss 1998b, 14.
43. Hardin 1968; and Stein 1983.
44. Underdal 1980.
45. Pauly et al. 1998; Myers and Worm 2003; Downs, Rocke, and Barsoom 1996; and Miles et al. 2002.
46. Young 1992.
47. Koremenos, Lipson, and Snidal 2001.

cal capacity to constrain access to the resource (unlike many coastal, lake, and river fisheries but also unlike restraints on the harvest of polar bear and tropical timber). High seas fisheries also, as a class, tend to be more symmetric in incentives and capacities than pollution problems which include, as a class, some symmetric and some asymmetric (upstream/downstream) externalities. Although illustrated using high seas fisheries, this process of winnowing cases could identify relatively large sets of other environmental problems that are sufficiently similar in problem structure to support comparative analysis, whether bilateral upstream/downstream pollution problems, multilateral regional seas agreements, or global marine pollution control agreements. The point here is that the value of using empirical distinctions to select cases depends on the degree to which they simultaneously reduce variation in theoretically-interesting aspects of problem structure. As theoretical consensus on, and operationalizations of, problem structure variation develops, we might hope to identify less obvious and more conceptual groupings of cases.

Within a set of nominally similar cases, opportunities may exist to further reduce variation in problem structure. Among high seas fisheries cases, exploitation incentives depend both on the demand for a species and on the capacity of relevant fleets. Focusing on agreements that regulate similar species in different parts of the world but from which a relatively similar international fleet supplies a global market seems likely to reduce variation in exploitation pressures that might otherwise explain significant differences in institutional outcomes. Equally important, scientific uncertainty about population dynamics, appropriate quotas, etc. varies across species and that variation is likely to influence the types and levels of regulatory restraints states accept and the degree to which they subsequently reduce their catch. Such species-specific aspects of problem structure seem likely to influence institutional design and observed behavior, suggesting the value of comparing agreements that regulate related species. Eight multilateral agreements manage high seas tuna fishing. Although two of these have never entered into force, protocols, amendments, regulations, and recommendations under the other six provide opportunities to examine the variety of techniques used to address this problem, in effect creating far more than six cases of study. Alternatively, one might assume that species-specific variation is not sufficiently influential that appropriate comparisons can be made among all, or some larger subset of, high seas fisheries. In either event, self-conscious case selection provides confidence that variation in problem structure among the cases studied is less than that among other sets of cases.

*Evaluating Problem Structure Variables and their Influence on Design and Behavior*

Even among sets of carefully-selected cases, variation is likely to remain in at least some aspects of problem structure. Identifying such variation ensures that it can be taken into account as a source of variation both in institutional design

and in behavioral outcomes. The obstacles to identifying variation in aspects of problem structure, noted above, are mitigated when making relative comparisons of the value of problem structure variables across a set of agreements selected to control for other aspects of problem structure. Thus, among tuna fisheries, which vary by both region and specific tuna species, it would be important to identify variation in the number and relative power of actors involved in each fishery (e.g., using harvest level and fleet size), the state of knowledge regarding the specific fish stock (e.g., using stock assessments), the pressures on each fishery (e.g., using world market prices), and some measure of changes in the technology of harvest (e.g., using average vessel size). Such explicit evaluation of variation in problem structure can determine whether case selection controlled for problem structure as intended and, even if it has, clarifies the value at which relevant variables have been controlled. It thereby identifies the conditions under which other variables are operating, providing the necessary foundation for the issues of interaction and endogeneity discussed above.

Endogeneity further dictates the need to examine how problem structure influences institutional design. Variation in institutional design among cases selected to control problem structure provides both a basis for evaluating hypotheses about how problem structure dictates institutional design and allows identification of the extent to which variation in institutional design remains possible within a given problem structure. Thus, among high seas tuna fisheries that involve collaboration problems, theory leads us to expect strong incentives to cheat and either carefully-crafted monitoring and enforcement provisions or an agreement that is ineffective because it lacks such provisions.[48] Given a goal of fostering continued harvest in the face of a dynamic population system, most fishery agreements must establish some form of ongoing adaptive management, usually involving intergovernmental commissions to make periodic regulatory decisions and scientific committees to inform those decisions.[49] The presence of scientific committees helps illustrate the endogeneity problem—their presence in all high seas fisheries agreements would suggest that the political exigencies of such problems preclude establishing agreements that lack such committees and therefore suggest comparing agreements to a counterfactual involving alternative scientific committee arrangements rather than to counterfactuals involving the complete absence of a scientific committee. Theory is less clear, however, with respect to the form of an institution's primary rules in response to a collaboration problem. In high seas tuna agreements, for example, states have, inter alia, established quotas, regulated minimum fish size, established closed seasons or areas, and limited effort and have done so using both binding and nonbinding approaches. Thus, in some cases, fundamental similarities in incentive

48. Stein 1983; and Downs, Rocke, and Barsoom 1996.
49. Young 1999b, 28–29.

structures leave states leeway in how to regulate a problem, and we might expect variation in regulatory approaches to lead to variation in institutional effectiveness.

Even when problem structure does not vary across agreements, it may well vary over time.[50] Uncertainty in particular is likely to decrease as scientific understanding improves.[51] Collecting evidence of how uncertainty has changed over time would shed light on claims that greater uncertainty produces institutions that are more centralized and that allow member states greater flexibility.[52] This point also illustrates the risks posed by endogeneity. Many fisheries agreements adopt more stringent regulations over time; if states subsequently show greater restraint in harvest, then the crucial question becomes whether that restraint reflects fishing fleets responding directly to improved knowledge about what harvest levels will maintain the fish stock, responding to the effects of that improved knowledge as mediated through the institutional devise of more stringent regulations, or the independent effect of more stringent regulations.

*Evaluating Effectiveness in Terms Appropriate to the Problem Structure*

Finally, as already noted, knowing that we are examining a set of tuna agreements allows us to specify more explicitly what constitutes institutional effectiveness. Catch data from the UN Food and Agriculture Organization or secretariats makes many, if not all, fisheries agreements potentially analyzable. Consider some available quantitative strategies to compare the effectiveness of fisheries agreement. An analysis might compare the average harvest or average change in harvest of states subject to an agreement's influence to an estimate of the harvest of those states, ceteris paribus, in the absence of such influence (either their own previous behavior or the behavior of states that never became members). Such an analysis might be appropriate for the many, though not all, pollution and wildlife conservation agreements in which states define the problem as requiring ongoing restraint in particular behaviors, e.g., eliminating ozone-depleting substances or banning whale hunting. In such cases, effectiveness can be inferred from levels or trends in the post-agreement period below corresponding counterfactuals predicted in the absence of the agreement. But such an approach ignores the evaluative implications of problem structure just discussed. Fisheries management agreements generally seek to improve long-term harvest levels by either sharply limiting harvest to allow populations to recover quickly or limiting harvest less sharply to help industry remain viable in the short to medium term. In both cases, however, institutional effectiveness in a "goal achievement" sense[53] requires evidence that post-agreement harvest lev-

50. See, for example, Underdal 2002b, 13; and Young 1999a.
51. Haas 1990.
52. Koremenos, Lipson, and Snidal 2001.
53. See Young and Levy 1999.

els were initially held below levels expected due to previous overexploitation but were later allowed to increase to sustainable levels above those expected due to such overexploitation. Evaluating such two-stage effectiveness might involve using prior knowledge regarding, for example, a species recovery rate to specify a date to generate two post-agreement periods for comparison. Alternatively, statistical techniques could be used to identify points of significant change in harvest data with the aim of assessing whether those structural breaks appear correlated with the establishment of, or changes in the design of, the agreement. Running identical econometric models on several fisheries agreements would generate statistics that could be used to compare agreement effectiveness. Quantitative analytic techniques are by no means necessary to this task—parallel qualitative strategies could work equally well. In short, selecting cases to reduce variation in problem structure allows identification of a problem-specific definition of effectiveness that applies well to the agreements studied but would apply less well to a broader range of agreements.

## Conclusion

Arild Underdal has provided scholars with an impressive set of theoretical, methodological, and empirical tools for investigating the effectiveness of international environmental agreements. He has provided invaluable insights and leadership in identifying which international environmental agreements perform better, which institutional design features explain the more successful ones, and the methods for doing both systematically and rigorously. This article has sought to build on Underdal's contributions by highlighting the necessity of taking problem structure seriously. Institutional analyses should account for aspects of problem structure as alternatives to, and as variables that interact with, institutional design in determining outcomes but should be particularly attentive to the endogeneity of institutional design to the problem structure-outcome relationship. The shortcomings of existing efforts have been highlighted to identify the obstacles that inhibit the incorporation of problem structure into institutional effectiveness studies and to suggest ways to overcome them. Confident and convincing arguments about which of several agreements were more effective and why requires that we develop clearer categorizations of the variation in problem structure, control such variation through careful case selection, investigate endogeneity in terms of how variation in problem structure dictates both institutional design and post-agreement behaviors, and evaluate effectiveness in terms appropriate to the underlying problem structure. Opportunities exist to respond to these analytic difficulties. But solving them requires considerably more effort and thought if we are to realize the promise of Underdal's intellectual contributions and contribute to policy efforts to address the numerous environmental challenges currently facing the globe.

## References

Brown Weiss, Edith, and Harold K. Jacobson, eds. 1998. *Engaging Countries: Strengthening Compliance with International Environmental Accords.* Cambridge, MA: MIT Press.

Chayes, Abram, and Antonia Handler Chayes. 1995. *The New Sovereignty: Compliance with International Regulatory Agreements.* Cambridge, MA: Harvard University Press.

Downs, George W., David M. Rocke, and Peter N. Barsoom. 1996. Is the Good News About Compliance Good News About Cooperation? *International Organization* 50 (3): 379–406.

Finnemore, Martha. 1996. *National Interests in International Society.* Ithaca, NY: Cornell University Press.

Haas, Peter M. 1990. *Saving the Mediterranean: The Politics of International Environmental Cooperation.* New York, NY: Columbia University Press.

Hardin, Garrett. 1968. The Tragedy of the Commons. *Science* 162 (3859): 1243–1248.

Hovi, Jon, Detlef F. Sprinz, and Arild Underdal. 2003. The Oslo-Potsdam Solution to Measuring Regime Effectiveness: Critique, Response, and the Road Ahead. *Global Environmental Politics* 3 (3): 74–96.

Jacobson, Harold K., and Edith Brown Weiss. 1998a. Assessing the Record and Designing Strategies to Engage Countries. In *Engaging Countries: Strengthening Compliance with International Environmental Accords,* edited by Edith Brown Weiss and Harold K. Jacobson. Cambridge, MA: MIT Press.

_____. 1998b. A Framework for Analysis. In *Engaging Countries: Strengthening Compliance with International Environmental Accords,* edited by Edith Brown Weiss and Harold K. Jacobson. Cambridge, MA: MIT Press.

Keohane, Robert O., and Lisa L. Martin. 2003. Institutional Theory as a Research Program. In *Progress in International Relations Theory: Appraising the Field,* edited by Colin Elman and Miriam Elman. Cambridge, MA: MIT Press.

Keohane, Robert O., Peter M. Haas, and Marc A. Levy. 1993. The Effectiveness of International Environmental Institutions. In *Institutions for the Earth: Sources of Effective International Environmental Protection,* edited by Robert O. Keohane and Marc A. Levy. Cambridge, MA: MIT Press.

Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. The Rational Design of International Institutions. *International Organization* 55 (4): 761–799.

_____, eds. 2004. *The Rational Design of International Institutions.* New York, NY: Cambridge University Press.

Martin, Lisa L. 1992. Interests, Power, and Multilateralism. *International Organization* 46 (4): 765–792.

Mearsheimer, John. 1995. The False Promise of International Institutions. *International Security* 19 (3): 5–49.

Miles, Edward L., Arild Underdal, Steinar Andresen, Jørgen Wettestad, Jon Birger Skjærseth, and Elaine M. Carlin, eds. 2002. *Environmental Regime Effectiveness: Confronting Theory with Evidence.* Cambridge, MA: MIT Press.

Mitchell, Ronald B. 2002. A Quantitative Approach to Evaluating International Environmental Regimes. *Global Environmental Politics* 2 (4): 58–83.

_____. 2003. International Environmental Agreements: A Survey of Their Features, Formation, and Effects. *Annual Review of Environment and Resources* 28: 429–461.

Mitchell, Ronald B., and Patricia Keilbach. 2001. Reciprocity, Coercion, or Exchange:

Symmetry, Asymmetry and Power in Institutional Design. *International Organization* 55 (4): 891–917.
Myers, Ransom A., and Boris Worm. 2003. Rapid Worldwide Depletion of Predatory Fish Communities. *Nature* 423 (6937): 280–283.
Pauly, Daniel, Villy Christensen, Johanne Dalsgaard, Rainer Froese, and Francisco Torres Jr. 1998. Fishing Down Marine Food Webs. *Science* 279: 860–863.
Rittberger, Volker, and Michael Zürn. 1990. Towards Regulated Anarchy in East-West Relations: Causes and Consequences of East-West Regimes. In *International Regimes in East-West Politics,* edited by Volker Rittberger. London: Pinter.
Simmons, Beth A., and Daniel J. Hopkins. 2005. The Constraining Power of International Treaties. *American Political Science Review* 99 (4): 623–631.
Sprinz, Detlef, and Carsten Helm. 1999. The Effect of Global Environmental Regimes: A Measurement Concept. *International Political Science Review* 20 (4): 359–369.
Stein, Arthur A. 1983. Coordination and Collaboration: Regimes in an Anarchic World. In *International Regimes,* edited by Stephen D. Krasner. Ithaca, NY: Cornell University Press.
Strange, Susan. 1983. Cave! Hic Dragones: A Critique of Regime Analysis. In *International Regimes,* edited by Stephen D. Krasner. Ithaca, NY: Cornell University Press.
Underdal, Arild. 1980. *The Politics of International Fisheries Management: The Case of the Northeast Atlantic.* Oslo: Universitetsforlaget.
_____. 1992. The Concept of Regime 'Effectiveness'. *Cooperation and Conflict* 27 (3): 227–240.
_____. 1998. Explaining Compliance and Defection: Three Models. *European Journal of International Relations* 4 (1): 5–30.
_____. 2002a. Conclusions: Patterns of Regime Effectiveness. In *Environmental Regime Effectiveness: Confronting Theory with Evidence,* edited by Edward L. Miles, Arild Underdal, Steinar Andresen, Jørgen Wettestad, Jon Birger Skjærseth and Elaine M. Carlin. Cambridge, MA: MIT Press.
_____. 2002b. One Question, Two Answers. In *Environmental Regime Effectiveness: Confronting Theory with Evidence,* edited by Edward L. Miles, Arild Underdal, Steinar Andresen, Jørgen Wettestad, Jon Birger Skjærseth and Elaine M. Carlin. Cambridge, MA: MIT Press.
_____. 2004. Methodological Challenges in the Study of Regime Effectiveness. In *Regime Consequences: Methodological Challenges and Research Strategies,* edited by Arild Underdal and Oran R. Young. Dordrecht: Kluwer Academic Publishers.
Underdal, Arild, and Oran R. Young, eds. 2004. *Regime Consequences: Methodological Challenges and Research Strategies.* Dordrecht: Kluwer Academic Publishers.
Victor, David G., Kal Raustiala, and Eugene B. Skolnikoff, eds. 1998. *The Implementation and Effectiveness of International Environmental Commitments.* Cambridge, MA: MIT Press.
von Stein, Jana. 2005. Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance. *American Political Science Review* 99 (4): 611–622.
Young, Oran R. 1992. The Effectiveness of International Institutions: Hard Cases and Critical Variables. In *Governance without Government: Change and Order in World Politics,* edited by James N. Rosenau and Ernst-Otto Czempiel. New York, NY: Cambridge University Press.
_____., ed. 1999a. *The Effectiveness of International Environmental Regimes: Causal Connections and Behavioral Mechanisms.* Cambridge, MA: MIT Press.

_____. 1999b. *Governance in World Affairs.* Ithaca, NY: Cornell University Press.
_____. 1999c. Regime Effectiveness: Taking Stock. In *Effectiveness of International Environmental Regimes: Causal Connections and Behavioral Mechanisms,* edited by Oran R. Young. Cambridge, MA: MIT Press.
_____. 2003. Determining Regime Effectiveness: A Commentary on the Oslo-Potsdam Solution. *Global Environmental Politics* 3 (3): 97–104.
Young, Oran R., and Marc A. Levy. 1999. The Effectiveness of International Environmental Regimes. In *The Effectiveness of International Environmental Regimes: Causal Connections and Behavioral Mechanisms,* edited by Oran R. Young. Cambridge, MA: MIT Press.