Rational Design: Looking Back to Move Forward

Barbara Koremenos, Charles Lipson, and Duncan Snidal

Why are international institutions organized in such different ways? Some, like the World Trade Organization (WTO) and the World Health Organization, seek very wide memberships. Others, like the Organization for Economic Cooperation and Development and the Group of Eight, are deliberately restricted. Some, like the UN, cover an extremely broad range of issues. Others are narrowly focused, dealing with a single product (Organization of Petroleum Exporting Countries, OPEC) or a single problem (Convention on International Trade in Endangered Species). Some, like the International Monetary Fund (IMF), perform a variety of centralized tasks and even negotiate sensitive economic policies with member states. Others do little more than organize meetings and collate information, as the Asia-Pacific Economic Cooperation forum does for its members. Most institutions allocate votes equally to all members. But a few of the more important institutions—including the IMF, European Union (EU), and UN Security Council—give large members more votes and effective veto power. Some institutions, like the Outer Space Treaty, are built around rigid promises. Others, like the WTO, allow states to alter their obligations when faced with unusual circumstances.

All these institutions address serious problems of international cooperation, but they are designed in very different ways to cope with them. What explains these differences in institutional design? The Rational Design project has one overriding aim: *to make explicit the connections between specific cooperation problems and their institutional solutions.*[1] To transform this broad goal into a workable research

We thank Jeffrey Smith, Marc Trachtenberg, Deborah Larson, Matthew Baum, Stanley Sheinbaum, the anonymous reviewers, and the editors of *IO*, David Lake and Peter Gourevitch.

1. In the volume's introduction we define *international institution* broadly enough to include treaties but not so broadly as to include tacit bargaining; Koremenos, Lipson, and Snidal, this volume.

International Organization 55, 4, Autumn 2001, pp. 1051–1082

program, we decomposed the idea of "design" into five measurable dimensions where institutions vary markedly:

MEMBERSHIP	Who is included and excluded?
SCOPE	Which issues are covered?
CENTRALIZATION	How centralized are the main tasks?
CONTROL	Who exerts control within the institution?
FLEXIBILITY	How are changing circumstances accommodated?

The contributors to this special issue of IO have used these dimensions in their articles, allowing us to compare and aggregate the findings. To explain variation in institutional designs, we focused on four independent variables often used to explain the success or failure of cooperation: enforcement and distribution problems, the number of actors involved, and several types of uncertainty. We drew on rational-choice theory to develop a series of testable and refutable conjectures that explain the design characteristics of international institutions in terms of these independent variables; we presented these conjectures in the volume's introduction. Subsequent articles evaluated these conjectures in specific areas of military security, the global economy, and the environment. Finally, Alexander Wendt offered an outside response to these efforts. In this article we provide an overview of the results and discuss the research agenda that lies ahead.

The main aim of the Rational Design project was to develop an explanatory framework and begin to test it against available empirical evidence. We asked the contributors of the case studies to evaluate any and all of our conjectures that related to their cases and to introduce other explanatory arguments as they felt necessary. We did not ask them to evaluate our conjectures against competing theoretical explanations (such as those drawn from realism or constructivism). Indeed, we intentionally departed from the perspective adopted by Alexander Wendt that argues for the need to pit theories against one another. While such "three-cornered fights" have their place, they are neither necessary nor desirable.[2] They are unnecessary because conjectures like ours can be tested against empirical evidence without reference to other theories. Our approach is appropriate because our goal is to understand the world, not to win a hollow victory among theories.[3] Moreover, contests among theories tend to exaggerate the antagonisms among approaches while slighting their complementarities. Indeed, much IR research has implicitly endorsed an erroneous presumption that an argument can only be shown to be right by showing that an alternative argument is wrong. This misses the point: all of our

2. Van Evera 1997.

3. We had pragmatic reasons for sticking to a "two-cornered" fight between our conjectures and the evidence. First, we did not want to expend the energies of our contributors on "meta-" and "-ology" debates; we asked them instead to focus on what the rational perspective could accomplish. Second, we did not feel competent to present, for example, the constructivist or realist conjectures about institutional design. If they "lost" such a fight, we would be accused of rigging the contest through our incompetence. While we welcome third parties to join this intellectual effort, we argue below that there is more payoff from doing so as allies than as adversaries.

arguments are incomplete and the alternatives may explain different aspects of the same problem. A more constructive approach is to evaluate each theory on its own terms and to take alternatives seriously by asking what they can explain that others cannot. We adopt this strategy here by using other approaches as guides to the deficiencies of our project and, in a speculative manner, asking whether these defects can be remedied within our framework.

Ours is not the only possible rational-choice approach to institutional design.[4] We focus on actors—usually, but not necessarily, states—trying to solve joint problems at the international level.[5] We emphasize particular problems that play a central role in the rational-choice literature, such as uncertainty, free riding, compliance, and distribution struggles. Alternative rationalist approaches might emphasize actors with different motivations—such as maximizing domestic electoral support or providing "rents" to key groups—and focus on different explanatory variables.[6] Moreover, there is substantial literature in IR rational choice that addresses related points regarding how to make actor preferences endogenous, whether to treat institutions in static or dynamic terms, and what level of aggregation to employ, among others.[7] We do not encompass all of these diverse rational-choice approaches; although, again, we speculate whether some of their insights might be used to improve our analysis. Ours is a rational design framework, not the framework.

Within the rubric of rational choice, we have tried to ensure breadth. Some studies are abstract and deductive; others are more detailed and empirical. Some deal with security issues, others with economic or environmental ones. In principle our conjectures should cover all types of international issues and institutions and all types of international actors. The variety of cases included in this volume is a step in that direction.

We begin with an overview of the findings and an assessment of how the Rational Design conjectures fared. We consider the results on each dependent variable and briefly discuss the independent variables. The results are generally positive, although in a few areas they are mixed or inconclusive. We then turn to the broader implications of our findings, concentrating on several topics directly related to

4. We thank an anonymous reviewer for reminding us of this point.

5. The Rational Design approach can include nonstate actors and does not inherently restrict the autonomous capacity of international institutions. Multinational firms and transnational interest groups can and do act at the international level and sometimes shape specific institutions, as Mattli makes clear. The Rational Design approach can include any type of intentional actor. In practice, however, states shape nearly all important international institutions, and for simplicity we sometimes limit our analysis to them. We also recognize that states can deliberately choose to endow international institutions with more autonomous powers to serve their purposes. According to Andrew Moravcsik, the principal members of the EU have done exactly that. Moravcsik 1998.

6. According to the influential work of George Stigler, Mancur Olson, Douglass North, and others, rent seeking is central to the organization of the state and state-society relations. Their insights about domestic politics could be extended to international cartels, including producer organizations like OPEC, and perhaps to a wider range of international institutions.

7. Many of these issues are taken up in Lake and Powell 1999.

institutional design and its systematic study. We first consider the level of formalization. The main issues here are the trade-offs involved in creating fully deductive models. Second, we conclude that "power" needs to be analyzed more fully and explicitly. Our discussion of control is at least a step in that direction. We also raise questions about how domestic politics affects international institutions. Domestic politics is auxiliary to the Rational Design framework, but we have tried to ensure that, when it is invoked, it is not used in ad hoc ways or as an after-the-fact explanation. Finally, we begin a dialog about how and why institutions change, particularly how they respond to changing preferences, external shocks, and other factors.

Empirical Results

The Rational Design project moves us beyond the question of whether institutions promote cooperation to an analysis of how specific institutional features promote cooperation. By providing an analysis of the importance of institutional design, the project also begins to paint a more detailed picture of how institutions do their work. Before we turn to the specifics, however, three general observations about international institutions stand out in our empirical results.

First, centralization is important. It enables actors to manage the different types of uncertainty they face and eases their cooperation problems when numbers are large. However, centralization is not a matter of "strength" or even of formalization. Instead, it reflects the important roles that central actors play in coordination and communication among participants.

Second, membership is important but operates differently than we anticipated. Its impact does not come by operating as a form of exclusion to address the long-standing IR concern for enforcement; instead, its primary impact is through softer (though not less important) roles in providing information and assurance.

Third, flexibility is one of the most important design features of international institutions. It provides a primary way to deal with the pervasive uncertainty in international politics, as well as a means to address distributional issues.

With this overview in mind, we now consider the specific findings on the conjectures and institutional design variables.

Summary of Results

Table 1 brings together the conjectures and the results of the eight case studies presented in this volume. Not every case study was relevant to every conjecture; we asked contributors to focus only on those conjectures their research spoke to most directly, positively or negatively. Indeed, one of the scope conjectures (S1, SCOPE increases with NUMBER) was not addressed by any of the studies. We also asked the contributors to pay particular attention to conjectures where their cases raised negative findings or where the causal mechanisms differed from those discussed

TABLE 1. Summary of results

Conjecture	Mitchell and Keilbach	Mattli	Oatley	Pahre	Richards	Rosendorff and Milner	Kydd	Morrow
M1: Restrictive MEMBERSHIP increases with the severity of the ENFORCEMENT problem	(−)				(−)			−
M2: Restrictive MEMBERSHIP increases with UNCERTAINTY ABOUT PREFERENCES							+	+
M3: MEMBERSHIP increases with the severity of the DISTRIBUTION problem				+	+			(+)
S1: SCOPE increases with NUMBER								
S2: SCOPE increases with the severity of the DISTRIBUTION problem	+				(−)			
S3: SCOPE increases with the severity of the ENFORCEMENT problem	+							
C1: CENTRALIZATION increases with UNCERTAINTY ABOUT BEHAVIOR		+			(−)			+
C2: CENTRALIZATION increases with UNCERTAINTY ABOUT THE STATE OF THE WORLD		+	+		(−)			
C3: CENTRALIZATION increases with NUMBER				+				+
C4: CENTRALIZATION increases with the severity of the ENFORCEMENT problem	(+)	+	+					−
V1: CONTROL decreases with NUMBER					−			
V2: Asymmetry of CONTROL increases with asymmetry of contributors (NUMBER)			(+)					
V3: CONTROL increases with UNCERTAINTY ABOUT THE STATE OF THE WORLD			+		+			
F1: FLEXIBILITY increases with UNCERTAINTY ABOUT THE STATE OF THE WORLD		+	+		+	+		+
F2: FLEXIBILITY increases with the severity of the DISTRIBUTION problem			+		+	+		−
F3: FLEXIBILITY decreases with NUMBER								+

Note: For full statements of the conjectures, see the introductory article.
Key: Conjecture is supported, +; supported weakly or with qualification, (+); not supported with qualification, (−); not supported, −.

under the conjectures. Thus the "silence" of the empty cells in the table does not indicate missing negative evidence but simply that certain elements of institutional design were either irrelevant to, or outside the focus of, the case study.[8]

8. These silences do raise another issue, however: the potential for selection bias. Results from cases where both the independent variable (X) and the dependent institutional variable (Y) are present might be overinterpreted as suggesting a "sufficient" condition that "X leads to Y." But a "silence" indicates that

In general, the results strongly support the Rational Design conjectures. Nearly 70 percent of the findings are strongly positive, and about 11 percent are strongly negative. (The remaining findings leaned one way or the other and often suggested a modification of the framework.) However, the strength of the results varied across institutional design features, so we consider the implications of the findings on different institutional dependent variables in turn. While we are obviously interested in whether the results speak to the Rational Design framework, we are equally interested in the substantive implications of the results.

Membership

While the overall results are positive, the conjectures on membership held a major surprise for our theoretical expectations. Two of the conjectures—M2, restrictive MEMBERSHIP increases with UNCERTAINTY ABOUT PREFERENCES, and M3, MEMBERSHIP increases with DISTRIBUTION problems—were strongly supported in two of the cases, with no significant contrary evidence. But conjecture M1, restrictive MEMBERSHIP increases with ENFORCEMENT problems, was contradicted in all three cases that addressed it.

We had expected membership to play a significant role in enforcement by operating as a mechanism for excluding noncooperators. This emphasis on membership as an exclusion device followed from our reliance on the logic of public goods that has been prominent in IR theory since hegemonic stability theory. On reflection, our failure to find support for this expectation is reinforced by the stylized observation that states are rarely expelled from international institutional arrangements. This renders enforcement through membership largely irrelevant since deviators are apparently rarely punished in this way.

Instead, the case studies found somewhat different membership mechanisms at play. Ronald Mitchell and Patricia Keilbach find that restricting membership will not help solve the enforcement problem in situations of asymmetric externalities because it misses the whole point of inducing perpetrators to join. If German industry is dumping chemicals in the Ruhr River and the Dutch are suffering downstream, keeping Germany out of any institutional solution hardly solves the problem. Andrew Kydd shows how membership operates as both an informational and an assurance device. Membership criteria not only provide useful signals for selecting good alliance partners but also have a significant impact on levels of trust with third party states. Finally, the support in James Morrow's study for conjecture M3, inclusive MEMBERSHIP increases with the severity of the DISTRIBUTION problem,

X can be present without Y necessarily occurring, so the "sufficiency" claim would be wrong. Ideally, we need to compare across cases to understand what other factors determine when X has the hypothesized effect Y, and when it does not. Although we are not aware of any situations where this problem arises, any interpretation of these results needs to be appropriately bounded in this way. Even with silences, however, a preponderance of positive findings still tells us that a factor does influence institutional design under some circumstances.

requires modification of the argument. Inclusive membership solves the distribution problem in prisoners of war (POW) treaties not by increasing possible trade-offs among a larger number of states but by signaling states' agreement to a set of rules and standards, the heart of the distribution problem. What we have learned from these proposed modifications is that sometimes there are alternative logics of institutional design and that it is essential to pay attention to the details of empirical cases.

These revisions highlight both opportunity and danger for the Rational Design approach. The opportunity is that this approach can handle a wide range of problems and we need not restrict ourselves to public goods or any other model. Reformulating the framework allows it to capture different contexts and nuances and lets empirical anomalies set the stage for progressive theoretical development. The danger, well understood by critics of rational choice, is that models will be modified to fit the data. Our response is to acknowledge the hazards (which are not unique to rational choice) and to emphasize that such "findings" need to be treated more in an exploratory mode than in a confirmatory mode. If the revised conjecture fits across multiple cases, however, we would then argue that it is beginning to acquire significant empirical support.[9]

Finally, since membership questions have been central to discussion in international political economy of multilateralism versus bilateralism and globalism versus regionalism, the relative paucity of findings in the political economy articles is surprising. Instead of being a disapproving silence, however, it reflects the contributors' focus on different design problems. Consider the two articles that do not discuss membership at all. Walter Mattli is concerned with why private actors choose one arbitration forum over another. Limitations on who participates are simply not relevant design features in his analysis. The same is true for Peter Rosendorff and Helen Milner as they develop a two-actor model to examine escape clauses. But to extend their model to larger numbers might require supplementary design features, including membership restrictions. How serious is the noise problem when larger numbers are using escape clauses? How credible are enforcement strategies in large groupings? In other words, would the institutional arrangements be renegotiation proof? The logic of Rosendorff and Milner's model agrees with our membership conjectures, but these findings still need to be worked out in detail.

9. Finally, we stress that the Rational Design conjectures were not retrofitted to the findings. We developed the framework before the case studies began. We subsequently refined our terminology, dropped one conjecture where we found its derivation poorly grounded, and dropped another where we found its logic to be identical to one of our other conjectures. Otherwise the conjectures were all derived before the evaluations. The conjecture we dropped on theoretical grounds dealt with the effects of distribution on centralization and had weak empirical support. The other addressed the effect of distribution on control and had strong empirical support. Dropping these conjectures does not affect the overall results.

Scope

The findings strongly support conjecture S2, SCOPE increases with DISTRIBUTION problems, and conjecture S3, SCOPE increases with ENFORCEMENT problems. In their detailed and thorough environmental study, Mitchell and Keilbach find very strong support for these two conjectures. Furthermore, they offer significant extensions of these conjectures by incorporating the impact of symmetry and asymmetry into the analysis and by differentiating positive and negative linkage. For example, when some states pollute and others suffer the consequences, it is impossible to win general agreement for institutions that rely heavily on sanctions. Polluters know that such arrangements will inevitably sanction them and provide few countervailing benefits. Not surprisingly, they try to block them or refuse to join. However, downstream states have an alternative. They can provide side payments to polluters for curbing their ways. This approach has problems of its own—it is rife with collective-action problems and perhaps adverse selection—but it does promise gains for both sides. The actors' heterogeneity leads to a particular institutional design, one that widens issue scope by rewarding good behavior. This potentially opens up wider questions regarding interesting ways to consider facets of "power" within the Rational Design framework. However, though these results confirm the conjectures, they do not resolve questions about generalizability beyond the specific case.

We list John Richards' finding regarding scope in the aviation issue as a qualified "negative," although it could be read either way. He argues that the scope of the Bermuda aviation institutions was determined by states' need to encompass all aspects of aviation competition. This accomplished, aviation already includes what Richards labels extensive "within issue-area" scope, so further external linkage is unnecessary. This recalls the long-standing problem in IR theory of dealing with issue linkage: the definition of issue-areas is itself endogenous.

Centralization

The centralization conjectures received extremely strong support, which draws attention to our conceptualization of centralization. We do not equate centralization with "formal" international organizations. Organizations with headquarters may simply be meeting forums for states that may not delegate anything at all. The International Labor Organization seems to fit this description: In one sense, it is a highly formalized institution, but with respect to substantive tasks performed, it is characterized by relatively low levels of centralization.

The negative findings, though relatively few, provide clues for how to broaden our understanding of centralization. Two of the negative findings come from Richards' analysis of international aviation markets. He argues that these findings are due to an unusual design interaction between centralization and control. Part of the problem in this case study lies in the difficulty of determining empirically whether the Bermuda institutions are centralized. For example, most important

decisions—such as routes, entry/exit, and number of passengers—were set by the bilateral agreements and therefore were not centralized. But the International Air Transport Association (IATA) did have centralized fare-setting ability. In particular, the IATA required that all countries unanimously set fares on major routes (both within and between regions), and fares for all smaller routes were based on the agreed price of the major routes. The IATA also added limited centralized monitoring and enforcement in the form of fare inspectors who monitored compliance with agreed-upon fares and issued fines for noncompliance. Despite these centralized mechanisms, Richards argues that the Bermuda institutions substitute control for centralization to overcome problems of uncertainty.

The other negative result is Morrow's finding on conjecture C4, CENTRALIZATION increases with ENFORCEMENT problems. In this case, the source of the enforcement problem lies in the two-level domestic problem between a state and its soldiers. Because the rules for handling POWs must be implemented on the battlefield, national authorities must be able to monitor and restrain their soldiers on the field. Again, Morrow uses an alternative rationalist logic—one that focuses on domestic actors rather than states as the key actors.

Mattli provides confirming evidence for three of the conjectures. For example, he conceives of centralization as an institutional arrangement that coordinates the interaction of two actors, sets procedures, collects information, and serves as the final arbiter. These roles make centralization very useful for resolving both enforcement and uncertainty problems. Enforcement problems arise when firms are no longer going to interact with each other—that is, there is no shadow of the future. Mattli argues that a centralized institution provides procedural safeguards and monitoring to overcome defection problems (conjecture C4). Uncertainty about the state of the world poses another major problem aggravated by the fact that companies are from different areas and their principals often speak different languages. Institutionalized arbitration procedures reduce uncertainty about what rules apply and increase understanding of those rules (conjecture C2, CENTRALIZATION increases with UNCERTAINTY ABOUT THE STATE OF THE WORLD).

One interesting finding revolves around centralization and states' reluctance to delegate power to an international organization. There are two positive findings for conjecture C4, CENTRALIZATION increases with the severity of the ENFORCEMENT problem. However, in neither case are states delegating to an international organization as the focal actor. Mattli's case involves private firms, not states. For Thomas Oatley, the focal actor is essentially the United States. Having a powerful state rather than an international organization play the role of the focal actor is also common in treaty guarantees. Consider the role of the United States in Middle East peace accords.

It is also important to note that centralization of one function does not necessarily serve as a platform for further centralization. Robert Pahre's treatment of "clustered negotiations" in nineteenth-century trade is a good example. Similarly, in the postwar trade regime, bargaining in the General Agreement on Tariffs and Trade (GATT) was clustered, or centralized, for decades before the partners decided to

create a more comprehensive institution with dispute panels, a larger staff, and some agenda-setting powers. In the case of POWs, as Morrow shows, states never sought to move beyond multilateral bargaining and the establishment of common rules.

Because our conception of centralization is very broad, an important avenue of inquiry will be to refine this concept into its different components. Centralized tasks may be as mundane as organizing meetings and assembling information provided by member states. But states may also delegate "stronger" centralized tasks, which range from intrusive monitoring to the orchestration of international enforcement. This finer differentiation of tasks is only implicit in our conjectures, and its analysis begs a series of important empirical and theoretical questions.

Finally, contrary to the common presumption that centralized institutions are more effective, more centralization is not always better. States design international institutions to meet specific needs, and decentralized arrangements often serve this purpose more effectively.

Control

The findings on control, though mainly positive, are sparse, so we do not claim too much for them. The two positive findings on conjecture V3, CONTROL increases with UNCERTAINTY ABOUT THE STATE OF THE WORLD, reinforce our general finding that international institutions are important in managing uncertainty—a point we return to later. Oatley's case provides modest support for conjecture V2, asymmetry of CONTROL increases with asymmetry of contributors (NUMBER). The decision-making rule of the European Payments Union (EPU) reflects a compromise between U.S. and European interests. As the source of hard currency, the United States wanted the EPU to exercise substantial control over European governments' macroeconomic policies. The United States established majority rule decision making in the EPU's managing board and wielded additional influence through U.S.-sponsored conditional aid programs. European members, however, had concerns about distribution and uncertainty. These concerns were accommodated by circumscribing the managing board's authority and by providing each European government with a veto over the managing board's decisions in the Organization for European Economic Cooperation. The EPU example illustrates another common feature of control rules: They may need to be quite complex to balance the competing needs of different members.

The one negative result, that states did not give up individual control as numbers increased in the aviation regime, reflects the very high sovereignty costs involved in air transportation during that period for both symbolic and security purposes.

Flexibility

The conjectures on flexibility are very strongly supported. This is gratifying because flexibility as a dimension of institutionalization has not been studied extensively. Our results show that it is a pervasive and important feature of international

institutional design—and that the rational approach appears to provide significant insights into it.

If anything, the empirical elaborations show that flexibility is even more important than we anticipated theoretically. Oatley's analysis of European trade and payments after World War II, Richards' analysis of the IATA, and Rosendorff and Milner's analysis of trade agreements all illustrate how flexibility can be designed to permit adjustment within existing institutional arrangements. Provisions in each case allow actors to respond to unanticipated changes in economic circumstances without dismantling the institution. Incorporating such provisions insulates the institutions from shocks that could otherwise threaten their existence.

Flexibility can also be important if actors learn over time about new technologies and tasks. With respect to arbitration rules in rapidly changing industries, Mattli argues that "firms operating at the forefront of new production and exchange methods are likely to prefer a flexible form of dispute resolution."

Importantly, the contributors suggest an expanded definition of flexibility. Oatley argues for one where the rules themselves allow some slack in what would otherwise be a binding constraint. Oatley deals with trade and money, but Morrow reaches similar conclusions about POW rules. His argument is very similar to recent attention given to "soft law" mechanisms that allow states to gain many of the advantages of legalization without being too tightly bound by them.[10] Because the treatment of POWs is so difficult to assess, and responsibility for violations so hard to pin down, some flexibility is essential to avoid mistaken punishments. Flexibility, in other words, helps actors deal with "noise" in the system.[11] If every perceived violation (including nonexistent ones) triggered a response, tit-for-tat feuds would be constant. Rational actors, well aware of the problems arising from erroneous sanctions, have good reasons to design a system with some flexibility.[12] Morrow's analysis suggests an intriguing relationship between flexibility and uncertainty about behavior, a relationship we did not examine in the volume's introduction.

Moreover, while we coded Morrow's finding regarding the relationship between flexibility and distribution (conjecture F2) as negative, it is worth noting that he does indeed uncover some flexibility in the POW treaties that helps solve the distribution problem as he defines it. Importantly, states can file reservations to the treaty at the time of ratification. And, as Morrow states, "they can also make clarifying statements about how they interpret parts of the treaty or object to the membership of another state in the treaty at the time of ratification." Reservations and clarifying statements are forms of flexibility that were not discussed theoretically in the introduction.

10. Abbott and Snidal 2000.

11. Morrow defines noise as "the inability to determine exactly what happened and so to determine precisely the other side's responsibility for the event."

12. The rationale for this kind of flexibility can be found in Downs and Rocke 1990; they argue that in the face of uncertainty about behavior the optimal strategies are often those that specify thresholds that trigger retaliation.

Wendt's comments about flexibility go to the root of the friendly debate in which he engages the Rational Design approach. He proposes that in certain issue-areas, such as human rights or perhaps the environment, "logics of appropriateness" dominate so that states will prefer rigid criteria to govern their behavior.[13] In support, Wendt notes that "there is no exemption for mass murder in the human rights regime."[14] While that particular dramatic example is certainly true, human rights agreements in general are actually much more likely than economic agreements to incorporate escape clauses (24 percent versus 5.4 percent).[15] Thus the preliminary evidence suggests that our hypotheses will also apply to the human rights area.

Still, while we are heartened by the results regarding flexibility, Mitchell and Keilbach are wise to point out that rigidity has benefits as well, which the conjectures on flexibility fail to address. As they argue, states "will eschew such flexibility and accept more binding, specific rules if, as in asymmetric externalities, each side's institutional benefits depend critically on the other side's carrying out the exact terms of the agreed-upon exchange."

Design Alternatives

Design alternatives are important from both a theoretical and an empirical point of view. According to the conjectures, for example, distribution problems can be addressed by membership, scope, and flexibility. Theoretically, we do not offer any arguments that explain when one approach would be used in place of another or when different design combinations might be chosen. It is quite possible that substitutability among the design alternatives weakens our empirical results. For example, if we find evidence for one conjecture (such as C3, CENTRALIZATION increases with NUMBER), we cannot assume that another conjecture addressing the same problem (such as S1, SCOPE increases with NUMBER) would necessarily hold. The (hypothetical) choice of centralization over scope may reflect the differential costs of the alternative design solutions, or it could be the result of reserving scope to address some other problems such as distribution.

Generalizing about design interactions is difficult. Richards finds that states did not create centralized institutions in response to the uncertainty about the state of the world prevalent in aviation markets (conjecture C2). Instead, they relied on strong control mechanisms (conjecture V3). Richards is silent on conjecture C2 because V3 was used to solve the problem. Morrow, however, finds negative results for both conjecture M1, restrictive MEMBERSHIP increases with ENFORCEMENT problems, and

13. Wendt, this volume, 1028.
14. Ibid.
15. See Koremenos 2000, which analyzes a random sample of 149 international agreements. Similar results on the flexibility of human rights agreements also emerge when we look at renegotiation provisions, withdrawal clauses, and amendment provisions.

conjecture C4, CENTRALIZATION increases with ENFORCEMENT problems. Since both conjectures address enforcement problems, the logic of substitutability does not seem to be at play.[16]

Independent Variables

Although our primary interest is with the dependent variables of institutional design, some of the "findings" regarding the importance of different independent variables may also provide guidance for future research. Two of the "traditional" explanatory variables in IR—enforcement and number—play only relatively modest roles in the cases. Number appears in only five "tests" of a conjecture in the various cases, although four of those are "confirmations"; enforcement appears eight times, but in half of those cases the conjecture is disconfirmed. By contrast, distribution appears in nine tests, of which seven are confirmations. This provides strong support for the recent movement to bring distributional considerations to the foreground of international relations analysis.

Even more striking is that uncertainty, in its various forms, plays a major role across the conjectures. It appears in fourteen different tests, of which thirteen are confirmations. To be sure, this could be the result of our special interest in uncertainty as a variable leading us to develop more conjectures based on it. Uncertainty has always been central to the study of international politics, especially research guided by recent innovations in game theoretic models. It is instructive to see what role it can play in explaining institutional design and how well those results explain crucial institutional features. This success shows that abstract models can have a significant empirical payoff when translated properly.

Drawing on models also forced us to separate enforcement and uncertainty as independent variables. Our definition of an enforcement problem is therefore quite narrow: given the strategic structure of the cooperation problems, are there incentives for actors to cheat that impede the ability of repeated play to support cooperation? Our analytic choice yielded some important insights. Whereas IR scholarship has tended to clump many things under the rubric of "enforcement problems," our results show that many of these are more precisely characterized as problems of uncertainty. When we do not know what state X is doing, we fear it might be cheating; when we do not know state X's preferences, we fear that it prefers to cheat. Institutions are useful in ameliorating the "enforcement" problems caused by uncertainty, but to solve the enforcement problems as we define them requires changing incentives or even preferences. In no case do we find the latter; Mitchell and Keilbach provide one example of the former.

16. An additional problem arises because we do not present an exhaustive set of conjectures. Nor, as we state in the introduction, do the ones presented all flow from one carefully-specified model.

Two broader implications of our definition of uncertainty should be noted. First, we differentiated among types of uncertainty—about preferences, about behavior, and about the state of the world—and showed how a clearer specification of uncertainty is important for developing claims about specific institutional features. In general, whether specificity is achieved through more detailed "verbal" interpretations or through formal derivations, it is essential for the fine-grained institutional approach we advocate.

Second, Wendt correctly points out that we use the conventional conception of uncertainty as equivalent to risk (that is, a probability distribution over known outcomes). This does not address broader conceptions of uncertainty as the truly unknown, where the possible outcomes cannot be enumerated or their probabilities are unknown. The criticism is valid and reflects our decision to focus on dimensions of institutionalization where we could draw most productively on existing rational-choice theory. Although we recognize the danger of letting our analysis be defined in terms of available methodological tools, it would be a greater mistake to let problems we currently cannot solve prevent us from tackling ones we can.

Our more restricted view of uncertainty also makes substantive sense insofar as states do not create extensive institutions in the dark. Indeed, a key aspect of negotiating agreements is replacing uncertainty with shared understandings (or common knowledge) of the problem at hand. For example, the first years of negotiations about global warming centered on achieving such commonality, and institutional progress is closely tied to success on that front. Some aspects of the process of converging to common knowledge may lie outside rational choice, which suggests important complementarities with constructivist and other ideational approaches. How to handle these broader forms of uncertainty is an area of active research both among game theorists and among proponents of alternative decision-making models.[17]

Broader Implications

Level of Formalization

The Rational Design approach has been criticized as both "not formal enough" (by those committed to fully deductive models based on axioms) and "too abstract and formal" (by those who see the rationalist approach as yielding too few concrete, empirically relevant results). Although these complaints are contradictory, we believe each arises for good reasons that should be taken into account.

17. Fearon and Wendt point out that parallel critiques of rational choice in terms of bounded rationality (such as Simon's) have since been significantly clarified and narrowed by subsequent work on incomplete information games. See Fearon and Wendt 2001; and Simon 1957. Ongoing game-theoretic work in the evolutionary tradition is exploring the implications of limits to rationality—including ignorance about possible outcomes. Fudenberg and Levine 1998; and Rubinstein 1998.

A comprehensive and precise set of conjectures, formally derived from axioms, is an attractive if rarely achieved goal across the social sciences. Unfortunately, such general models of international institutions are well beyond the reach of current IR scholarship.[18] Our initial conjectures were guided by tightly derived models. But those tended to be partial models of fairly specific circumstances that we had to adapt in a looser fashion to "derive" more general predictions about international institutions. To meet higher standards of rigor, the authors of the formal articles in this volume had to carefully limit their subject matter and remain at a somewhat abstract level. These articles illustrate the possibilities for deepening the larger program of rational institutional design not by building a comprehensive, general theory but by tailoring specific yet connected theories within the larger Rational Design framework. Efforts to bring in domestic politics (discussed later) illustrate the possibilities for expanding our analysis by building on other rational approaches to institutional design; they also show the limitations of formal approaches.

Although a single deductive model is a bridge too far, we hope we have shown how progress can be made with a systematic but less strictly formal approach to the same questions. Our goal was to lay out a consistent and quite general analytic framework and then to derive, by reference to formally established results, a set of conjectures that could be tested against the observed range of institutional design. In doing so, we took seriously the often repeated (and sometimes accurate) charges that rational-choice approaches lack empirical content, have little new to say, and fail to illuminate real-world political processes and outcomes.[19] Our results show, however, that these shortcomings are not inherent to the rational-choice analysis of institutional design.

On the empirical side, we have already discussed the significant findings that came out of the project. While the case studies could surely have been completed without the Rational Design framework, it offered important guidance for the individual articles and, more importantly, provides a systematic way to bring the empirical results together. It enables us to make comparisons across the cases, to see what holds and what does not, and to understand what needs to be modified according to the particular circumstances of individual cases. The results further show that rational-choice empirics need not be ad hoc. The conjectures themselves were clearly falsifiable, and some were shown not to have much support. Proposed modifications to the theory were offered in the spirit of refining and improving its explanatory power, not to armor it against empirical disconfirmation. On the theoretical side, the conjectures are deeply indebted to existing models, and several articles show how the conjectures can be developed further through formal analysis. Rosendorff and Milner's analysis of escape clauses in trade agreements elaborates one of the conjectures on flexibility (F1, FLEXIBILITY increases with UNCERTAINTY ABOUT THE STATE OF THE WORLD) and shows how it can be derived in a somewhat

18. Shubik makes an almost identical argument regarding the application of game theory to explain economic activity. Shubik 1982.

19. Walt 1999.

different way. Their focus on a different type of uncertainty (about domestic politics) shows that the conjectures are more generalizable even as they provide additional depth on the mechanisms underlying the conjecture. This is a good example of how the abstraction of models can bring together different substantive arguments. It also represents a significant "deepening" of the theory by drawing on other rationalist accounts.

Likewise, Kydd's model begins with the basic insight that, in organizations like NATO, membership criteria serve as costly signals to filter out states uncommitted to cooperation. He examines what happens when a nonmember country sees the group as a potentially hostile alliance. This addresses the important policy debate about whether some countries—particularly those once part of the Soviet Union—should not be included in NATO because of the provocative effects on Russia. The formal analysis shows that the more restrictive the membership criteria are, the less provocative expansion is to the outside power. If the criteria are restrictive enough, expansion may actually be reassuring by signaling that the alliance is not expanding as much as it would if it had truly aggressive designs on the outside power. Thus the formal analysis takes us beyond the conventional wisdom to illuminate possibilities not commonly discussed in the policy debate over NATO enlargement.

Several of the empirically oriented articles provide useful feedback on how the framework could be modified to better explain the international experience. Mitchell and Keilbach expand the conceptual basis of scope by differentiating positive versus negative linkages and by considering the existence of asymmetries in externalities. Pahre introduces the notion of clustering as a specific type of centralization that is particularly associated with "most favored nation" status and increases with number of actors. These studies illustrate how a combination of theoretical and empirical results can provide valuable guidance for subsequent theoretical work in the Rational Design framework.

It might seem odd that our "case studies" include two articles that are primarily formal exercises working out the logic of Rational Design-style conjectures in different settings (those by Kydd and by Rosendorff and Milner). Pahre develops a formal analysis as a prelude to a statistical analysis. Yet the inclusion of even purely formal articles is natural for two reasons. First, the conjectures we offered in the introduction represent a beginning (or a continuation) of work in bringing the rational-choice approach to bear on the design of international institutions. As should now be clear, the rational-choice theory applicable to institutional design is rich and complex; to pretend to capture all of the theory's implications even in our large set of conjectures would be hubris on our part. Moreover, as both the empirical and theoretical case studies show, contextual variation matters greatly in determining the design of institutions. The analytic articles represent a first step in showing how the logic of rational design works itself out differently according to these variations.

Second, and perhaps more surprising to some readers, the analytic articles actually contain significant empirical content. Each models a specific problem that has been identified empirically as well as theoretically. Each begins with what are

sometimes called "stylized facts," which are widely used by modeling practitioners but often disparaged by empiricists.[20] But good stylized facts have a great deal of empirical content. As a summary, they are not unlike a regression line that conveys the general tendency in a set of data even though it need not pass through (and thus characterize with strict accuracy) any of the individual data points. While they do not capture the details of any single case, such stylized summaries often convey more general information about the set of cases than can any subset of cases. Insofar as models bring together important stylized facts, they provide a possible explanation with significant empirical content. Of course, these are not disconfirming tests in any way, but as guides to the development of better theoretical understandings they bring along substantially more empirical content than is often recognized.[21]

Finally, critics will raise concerns about fitting theory to cases. Those who actually build models will have fewer concerns on this count simply because they know how difficult it is to construct a reasonable model that fits any potential observation. If it were possible, the same would be even truer of verbal theory, since in principle verbal theory subsumes formal theory and much more. Moreover, the explicitness of the assumptions in formal analysis opens it to scrutiny, which makes it easier to weed out implausible assumptions and errors in logic. In fact, by bringing together a wide range of conjectures, the Rational Design framework imposes stronger consistency requirements across different analyses. For example, if different conjectures conflict, the differences must be justified and explained in terms of underlying assumptions.

Power

The fundamental realist critique, beyond its general skepticism of the importance of international institutions, is that institutionalist analyses neglect power.[22] We did not emphasize power in the development of our conjectures because the formal literature does not offer compelling results.[23] However, our conjectures on control do address power relationships within institutions and accord roughly with realist intuitions (in particular, conjecture V2, asymmetry of CONTROL increases with asymmetry of contributors (NUMBER), and to some extent conjecture V3, CONTROL increases with UNCERTAINTY ABOUT THE STATE OF THE WORLD). While we found it

20. A stylized fact is a summary statement of a circumstance or relationship that is observed across a wide range of situations and is taken either as an input into a modeling exercise or as something to be explained through it.

21. Pahre (pers. comm. with Koremenos and Snidal) offers the interesting observation that the more formal articles address only one or two of the conjectures, whereas the nonformal articles address between four and nine conjectures. This reflects the power of formal research in exploring the logic of a particular conjecture and the power of empirical cases in examining evidence on multiple conjectures within a case.

22. Mearsheimer 1994/95.

23. Power is notoriously difficult to conceptualize and theorize—both in the traditional and in the formal literature. An example of the latter is the attempt to treat patience (that is, a low discount rate) as a proxy for power in bargaining models, which we find unpersuasive from a substantive perspective.

surprising that the empirical support for these conjectures is limited, these conjectures nevertheless demonstrate the possibility of incorporating such considerations within the Rational Design framework.

Two articles illustrate alternative ways to bring power into the existing framework. Mitchell and Keilbach's refinement of the scope variable with respect to asymmetric externalities allows a natural extension to power. If the victim is less powerful than the perpetrator, coercion is precluded; hence rewards must be used. If the victim is more powerful, we would expect coercion to be used since it is less costly from the victim's point of view. Their Rhine River pollution example is consistent with this: The weaker victim state, the Netherlands, had to compensate the more powerful perpetrator state, France. Importantly, this analysis can be extended beyond the environmental realm to the security realm, where realists argue that power is most influential. Consider the asymmetric externalities between the United States and North Korea over the latter's "pollution" of the existing nonproliferation system.[24] In this case, the perpetrator is clearly the weaker state; hence, while both rewards and coercion would in principle be possible, we would expect to see coercion. Ultimately, the United States rewarded North Korea with the transfer of light-water reactors. Although this may seem surprising at first glance, closer analysis of the case explains this design choice. North Korea's near autarchy means it is not very vulnerable to coercive tactics such as economic sanctions. As Mitchell and Keilbach argue, "A state's power is thus relational and issue specific, dependent on . . . how vulnerable and sensitive other countries are to the resources it controls."[25] This points out the difficulty of operationalizing power as an explanatory variable.

Power may also play an important role in political economy cases, as Oatley shows with respect to the role of U.S. power in directly affecting the design of the European payments system. Alternatively, power might serve as an intervening variable. For example, during the Bretton Woods negotiations, the U.S. executive used its negotiating power to limit the maximum contribution to the IMF to what Congress would be willing to authorize, thereby ensuring that no other state would have greater control, given weighted voting.[26] This illustrates conjecture V2, asymmetry of CONTROL increases with asymmetry of contributors (NUMBER), that is, power, although the logic is slightly different from what we suggested in the framing article.

Thus nothing in the Rational Design framework precludes power from being incorporated as an additional independent variable. We would expect it to improve our understanding of important dimensions of institutions, especially but not only with respect to security issues. For example, membership decisions during the Cold War were heavily informed by the tight control held by Western powers, which preferred to keep the Communist states out of many international economic

24. This example is from Kim 1999.
25. Mitchell and Keilbach, this volume, 898.
26. This example draws on Wheelbarger 1999.

institutions for larger political and ideological reasons. Hence, although we did not emphasize realist power considerations in our framework, we view them as highly compatible with it, especially given that the basic logic of the realist perspective is rational. Thus we do not see the Rational Design project as resolving any of the so-called debates between neoliberal and realist perspectives. Instead, the framework provides an analytical way to bring together their different substantive presumptions in order to evaluate them empirically.

An alternative critique by those who emphasize power considerations is that power interacts in a complex way with other explanatory variables and its inclusion may thereby undermine the entire framework.[27] A simple way to think about the omission of power as an independent variable is through the statistical theory of omitted variables. Omitting an independent variable may bias the estimated impact of the remaining independent variables. But this bias arises only if the included variables are correlated with the omitted one.

In our framework, it is not at all clear that power should be correlated with our independent variables: number, the severity of the distribution problem, the severity of the enforcement problem, and the three kinds of uncertainty. For example, consider the three kinds of uncertainty. State of the world uncertainty tends to be a characteristic of technology or of the natural environment (such as the state of the world when negotiating about global warming) and hence not likely to be correlated with power. With respect to uncertainty about preferences or behavior, perhaps it could be argued that powerful states are better able to conceal information. Yet we also observe today that the most powerful state, the United States, is one of the most transparent. Hence, there is no reason to think that these independent variables would be correlated one way or another.

In any event, this is an issue appropriately addressed by empirical researchers seeking to apply our framework on a case-by-case basis. Our theoretical framework proceeds when all other factors, such as power, are being held constant.

Finally, the explanatory leverage of power may come into play in the substantive terms of the cooperative deal,[28] in other words, "who gets what."[29] Moreover, the goals of the institutions themselves may be affected by the powerful states. Given that we focus on institutions that are created, there may be a nontrivial selection bias operating here.

Domestic Politics

Our framework does not explicitly include—or exclude—domestic politics. The contributors were free to incorporate domestic political considerations to whatever

27. We thank an anonymous reviewer for bringing up this point.

28. This is not meant to denigrate our focus on design. Quite the contrary, design matters if the substantive provisions are going to come into effect and stay in effect.

29. Of course, one of our design features, scope, is substantive in nature. Indeed, the determination of scope may depend on the availability of powerful states to keep items off the agenda (agriculture out of GATT, for example) or to force them on the agenda (services in the WTO, for example).

extent they considered appropriate in explaining their cases. We did not force a particular model of domestic politics on them since there is no widely accepted model available in the literature, and developing one anew was beyond the scope of this project. We did stipulate that domestic politics not be introduced as an ad hoc or post hoc explanation. We also asked the contributors to capture the effects of domestic politics through the variables in our framework whenever possible.

At some level—the claims of structural realism notwithstanding—all international politics has domestic roots. From a rational-choice perspective, states are abstract entities whose preferences can only be understood as a (sometimes problematic) aggregation of other interests.[30] Thus the question is not whether to deal with these considerations but how best to deal with them. In some cases (Mitchell and Keilbach's examination of environmental institutions is a good example), the appropriate response is to follow the longstanding IR tradition that treats states as unitary actors. By treating states as having well-defined preferences over environmental issues, Mitchell and Keilbach produce a convincing explanation of institutional scope without introducing extraneous and complicated domestic political concerns. Kydd, on the other hand, goes part way toward unpacking the domestic "black box" through his discussion of democracy as a signal of benign preferences. But adding in richer domestic detail would complicate his analysis without adding to the explanation.

Among the remaining contributors who did introduce domestic variables into their analyses, we find two broad approaches. The first incorporates key aspects of domestic politics as exogenous factors affecting state preferences or constraints. The second goes part of the way toward endogenizing domestic politics or incorporating them into an overall model of institutional choice. This difference roughly corresponds to the difference between treating domestic factors as determined outside the model and including domestic politics as the second level in a two-level game.

The cases show several ways to incorporate exogenous domestic political factors into institutional design. Some studies highlight a few domestic attributes as invariant characteristics of the environment in which international institutions operate. In Mattli's study of commercial arbitration, for instance, domestic legal environments provide exogenous constraints on arbitral institutions. Both in designing institutions and contesting individual cases, the participants must consider domestic courts. It matters that domestic courts are more responsive to "monitored" arbitration and that courts in some countries, such as India and Singapore, sometimes overrule arbitral awards or block their enforcement.

Other studies incorporate gross domestic characteristics as exogenous explanatory variables. These characteristics affect domestic preferences and, thereby, the existence and design of international institutions. For example, according to Rich-

30. We view factors such as a state's position in the international economy or its position with respect to the international distribution of power as constraints and environmental conditions, not as preferences. Over time, of course, the environment may condition state preferences. The Rational Design conjectures, however, are comparative static hypotheses.

ards, domestic commercial interests explain why the United States wanted international aviation institutions to promote competition and the United Kingdom wanted tighter regulation. Because U.S. airlines were more efficient than British ones at the time, they would probably have been the big winners in a more competitive regime. Similarly, Morrow's analysis of the treatment of POWs is based on exogenous domestic preferences in the form of Japanese cultural views of prisoners, German views about Slavic and Western prisoners, and so forth. But it is important not to confuse the role of preferences with the systemic analysis that Morrow's study illustrates. His analysis does not require a theory of how the Germans, Japanese, Americans, or British acquired their preferences regarding POWs, only that they had them. Morrow's specification of interests then plays out in interesting ways within the rational-choice analytic framework, as they do in the other articles that adopt this approach to the incorporation of domestic factors.

An alternative way to think about state preferences is in terms of general goals, such as achieving efficiency or securing reelection or maintaining power more generally. Even approaches that appear to eschew domestic politics incorporate it implicitly this way. Mitchell and Keilbach, for instance, implicitly treat states as having efficiency goals without examining the goals of domestic interest groups more closely. Oatley's article on the payments system after World War II provides a rich depiction of how national preferences vary with domestic political and economic interests. European members had varying employment objectives, which played a central role in determining national preferences over different international payments schemes.

In other cases, domestic politics is conceived of as a source of exogenous shocks to international institutions. In Rosendorff and Milner's article, the designers of trade institutions worry that domestic politics might produce destabilizing shocks. They cope with that possibility by building escape clauses into the institution. The sources of the shocks are not themselves formally modeled but presumably result from changes in the relative economic importance or political influence of various domestic producer and consumer interests. Similarly, in Richards' analysis of how airline deregulation threatened international arrangements, the original "shock" of U.S. deregulation arose from purely domestic considerations.[31] But when the United States moved to deregulate airlines at home, it fundamentally undermined the Bermuda institutions that regulated air travel abroad.

The second broad approach partly endogenizes domestic politics by fleshing out its impact on state preferences. Both case studies of trade adopt this approach. Pahre draws on trade policy models to justify the preferences of the governments in his model. Specifically, he uses an electorally motivated model to derive government preferences for nonzero domestic tariffs and zero foreign tariffs. These preferences inform Pahre's three-state bargaining model from which he derives his main

31. Our categories are not intended to pigeonhole each article uniquely. For example, Richards' consideration of domestic factors fits both "gross characteristics" and "exogenous shocks."

theoretical contributions. Likewise, in Rosendorff and Milner's analysis, an informal model of domestic politics underpins their formal model. The government cares about both consumer welfare and firm profits, reflecting the incumbents' concern with reelection.

A blend of the two broad approaches introduces domestic actors into the first-level international game. The studies by Richards and Mattli illustrate this approach by including both states and firms in their international analysis. For simplicity, however, both studies omit any domestic analysis of how firms lobby national governments. In this intermediate case, no domestic bargain is being struck. Instead, the domestic actors have a direct role in striking international bargains, so they are international actors in their own right.

We draw three lessons regarding domestic politics from the articles in this volume. First, rational choice provides several alternative ways to incorporate domestic politics into the analysis of international institutional design. The second concerns the value of theoretical flexibility in incorporating domestic politics. Domestic politics needs to be explicitly addressed to understand some international institutional arrangements but not others. Moreover, when domestic politics is important, the role it plays varies substantially across issue-areas and may be best addressed in different ways. Finally, the analyses here illustrate how these considerations can be incorporated into rational-choice analyses of international relations without requiring a full theory of domestic politics.

Preferences and Preference Change

We did not impose any general assumptions about actors' preferences.[32] Instead, we asked contributors to specify preference configurations in terms of distribution and enforcement problems relevant to their cases. Taking preferences as stable and exogenous is both simple and powerful, which is why it has long been standard in rational-choice modeling. It also avoids the misuse of "revealed preferences," where the preferences themselves are inferred from the very phenomena they are supposed to explain.

Treating preferences as exogenous does not imply they are ad hoc or arbitrary. Several studies develop clear theoretical underpinnings for the preferences they propose. This is clearest in formal articles such as those on trade institutions (Pahre, Rosendorff and Milner), where government objectives are driven by a concern for domestic welfare, distributional effects, and corresponding electoral incentives. Indeed, all the articles specify actors' goals in advance, whether it is to improve the treatment of their captured soldiers (Morrow) or to improve their external monetary balance while achieving other domestic economic goals (Oatley). Such assumptions are typically based on any existing empirical and theoretical consensus in the field in combination with more focused knowledge of the issue at hand.

32. For an in-depth discussion of related issues, see Lake and Powell 1999, especially the introduction and the chapters by Peter Gourevitch and Jeffry Frieden.

This clear specification of actor preferences is one of the prime virtues of any rationalist framework. Doing so also invites serious inquiry into the correctness of alternative specifications—for example, does vote maximizing or the pursuit of economic efficiency provide a better explanation of trade policy?—as well as into how preferences change (either exogenously or endogenously) and how institutions might change correspondingly, a topic we consider in the next section.

The relationship between "bedrock" preferences and constraints provides a guide to an empirical analysis of preference change. Bedrock preferences are fundamentally stable preferences, but composite actors or actors operating under constraints may sometimes be usefully modeled as if they had preferences that depend on the particular decision context. Consider the two-level setting discussed earlier, where one level is diplomatic relations among states and the other involves domestic politics within each state. The impact of one level on another can be thought of in terms of the constraints that one level imposes on the other. In this two-level arrangement, the executive has to negotiate an international agreement subject to the constraint of domestic ratification. An alternative conceptualization is to treat the two levels as one—component parts of a unitary actor—and treat the outcome of domestic bargaining as a change in preferences. An extreme example would be a successful trade regime that strengthens free-trade interests and severely weakens protectionist forces.[33] The cumulation of this process would forge "trading states" with very different goals from their mercantilist predecessors.[34]

Viewed this way, Rosendorff and Milner's escape clauses can be interpreted as states designing institutions for the possibility that they will have different preferences (instead of the same preferences but different domestic constraints) at some point in the future. Similarly, we often speak of rich and poor countries having different preferences over some issue (for example, environmental quality), but this difference in preferences could alternatively be expressed in terms of different constraints (for example, levels of development). While it might be desirable to express all of these relationships in terms of bedrock preferences and different constraints, it is sometimes empirically useful to sacrifice some of that generality in favor of more context-specific preferences (implicitly incorporating constraints). This allows us to specify plausible preferences using empirical knowledge of specific cases and to identify preference "change" within the cases.

One example of an empirically identifiable preference change is when a state becomes democratic (or communist or fascist). The assumption is that we can impute different (international) preferences to domestic leaders operating under the different constraints imposed by alternative regime types. This allows us to trace the impact of preference changes but does not account for the change itself, which remains exogenous.

33. Rogowski 1989.
34. Rosecrance 1986.

A related source of change—here in beliefs rather than preferences—occurs within the actors as they learn how the world works and therefore shift their informational constraint. Institutional flexibility facilitates this type of learning by lowering the risks of agreement while states discover how an agreement works.[35] Reassurance in security affairs illustrates another type of learning: states develop insights into one another's preferences and likely behavior, as Kydd's analysis shows. When such learning occurs at the collective level, it can generate an alternative and superior equilibrium that states achieve partly by designing appropriate institutions. This perspective captures the understanding that security communities are, above all, groupings of states that have developed common expectations about one another's pacific behavior—expectations that then change their individual and hence collective behavior. In this sense, our framework captures the important idea that "anarchy is what states make of it" and shows how rational purpose and institutions play a central role in constructing alternative equilibria.[36]

Another way that preferences can change is when different actors (with different preferences) become involved. Membership rules, for example, determine which actors are included in an institution and thus which preferences must be accommodated in its design. As membership increases, the heterogeneity of the group will typically increase. Admitting nongovernmental organizations or other private actors, for example, can add valuable expertise and information, but it also adds new voices and new preferences. Moreover, while we have not addressed international organizations as actors, they clearly have that role in certain situations. Some, such as the EU, are major players that operate with substantial independence and authority—their preferences have a major impact on issue outcomes.

Even if the actors remain the same, their preferences may change endogenously. Participation within international institutions, for example, may alter the participants' goals and the trade-offs they make. Successful international economic relations typically increase states' interests in pursuing deeper forms of economic cooperation. The steady development of the GATT/WTO framework is a good illustration. So, too, is the construction of a single European currency. These changes are really in "derived" preferences, that is, preferences over actions conditional upon the constraints imposed by other actors' behaviors and by the institutional environment. Hence, bedrock preferences are constant—a hallmark assumption and limitation of the rational-choice approach.

This discussion of preferences illustrates possible limitations to the Rational Design framework, but these limitations do not diminish its value. Rather, they point out challenges for theory development and testing. In some cases, rational-choice theory will have to be augmented by auxiliary and even competing theories. One source will be other rationalist accounts of the more detailed processes and mechanisms of preference formation, as in our earlier discussion of domestic

35. Koremenos 2001a.
36. Wendt 1992.

politics. Another source will come from outside rationalist approaches as suggested by Wendt's discussion of the logic of appropriateness. One of the continuing debates about the logics of consequences and appropriateness concerns the extent to which these are truly independent logics—as we believe they are in part—versus the extent to which they are deeply intertwined. Surely one reason we have seen the emergence of a "nuclear taboo," of human rights regimes, and of a delegitimization of bribe-paying is because of their consequences. Yet these changes also reflect fundamental changes in preferences that go beyond such direct effects.

Dynamics of Institutional Change

Throughout this volume, we have treated institutions as components of equilibria. They are designed to address specific problems and are supported, in turn, by those who benefit from the resulting equilibrium. This approach allows us to understand why institutions have their specific features, and comparing across equilibria (comparative statics) provides insights into how institutions adapt to different circumstances. Although the strength of the empirical results supports both our strategy and our view of institutions, the equilibrium assumption only indirectly addresses questions about institutional change.

Equilibrium analysis can also provide stronger guidance for a more dynamic analysis. Understanding how the components of an equilibrium fit together helps identify the circumstances under which an equilibrium becomes unstable and when a new equilibrium will emerge.[37] When a change in one or more of the basic parameters of a game, such as the number of players or the degree to which players value the future, invalidates the equilibrium, then the players are in the same position as they were when they originally negotiated the equilibrium: they must establish a new equilibrium. Often change is not so dramatic; the old equilibrium continues to be sustainable while new equilibria emerge.[38] Whether or not states exploit resulting opportunities is at least partly a function of the original institutional design. For example, planned review conferences and opportunities for renegotiation may facilitate Pareto-superior shifts.[39] Importantly, however, once an institution is formed, a subset of players can use opportunities for equilibria shifts to exploit other players.[40] Again, the possibility of such redistributive moves depends on the original design—in particular, the control rules.

Generally speaking, change is most interesting when there are multiple equilibria, including potential ones that depend on new institutional arrangements. Since different equilibria may favor different actors, choosing among them is politically charged. More is at stake here than technical shifts induced by changing circum-

37. Calvert 1995.
38. Koremenos 2001b.
39. Koremenos 1999b.
40. See Knight 1992; and Gruber 2000.

stances. Rather, exogenous changes set the table for tough political bargaining over which institutional equilibrium to pursue.

Changing an institutional equilibrium is often difficult because those who benefit from current arrangements resist change. This conservative bias is reinforced by other factors. The first is transaction costs—a dead weight that lessens or even eliminates the gains from changing equilibria. Second, many international actors are risk-averse and so are reluctant to gamble on untested solutions. Finally, the institution viewed as an actor may have an interest in defending its associated equilibrium, even if that is inefficient. This tendency may be offset by the availability of institutional mechanisms that promote nonthreatening change within the institution. Ultimately, of course, such resistance is bounded by the ability of states to abandon inefficient institutions.

Institutional stability means that existing institutional arrangements are determined not simply by existing conditions (the focus of comparative static analysis) but also by the historical development of the institution ("path dependence"). Because existing institutions reflect their development, they differ significantly from what might be created de novo. An important example is that the voice of rising powers is likely to be underweighted, and the voice of former powers overweighted, in long-standing institutions compared to how they would be weighted in brand new institutions. Because of institutional "stickiness" and path dependence, institutional design will be hotly contested since the choices persist and are not easily undone.[41]

Institutional stickiness generates a "stability/instability" paradox of institutions. The stability of institutions is central to their value since it allows actors to pursue their objectives, including plans to update the institutions themselves. Indeed, if institutions were highly malleable, they would have no impact in solving the problems states face. Yet this stickiness also means that over time institutions can deviate significantly from the collective goals of states and not effectively address current needs. Thus the very stability that gives institutions their value also allows them to build up instability-generating pressure as external conditions change. Therefore, one of the greatest challenges in institutional design is to find the optimal trade-off between stickiness and flexibility. Here we can usefully distinguish between two kinds of institutional change, one aimed at creating new institutional equilibria, the other at sustaining an existing equilibrium in the face of changing external circumstances.[42]

Many historic junctures of international institutionalization have entailed substantial interstate deliberations in search of arrangements that can better states' interactions. This is especially apparent in this volume's more historically oriented studies. Oatley, for instance, traces institutional change from a network of bilateral

41. A highly instructive example is the long-standing difficulty that states have had in agreeing on televisions standards—even though all would benefit from a common standard—because of future implications for competitiveness. See Austin and Milner 2001; and, more generally, Abbott and Snidal 2001.

42. This echoes Krasner's change of regime and change within a regime. Krasner 1983.

trade agreements into the multilateral European Payments Union. Similarly, Richards analyzes how the development of the Bermuda institutions was driven by change in aircraft technology and postwar market characteristics and suggests that the regime's breakdown was the result of changing state interests. Change is also relevant in the more formal articles, as exemplified by Kydd's discussion of the importance of the dynamics of reassurance to the expansion of NATO. Such institutional innovations begin with a clear understanding that it is impossible to predict all future contingencies and that outcomes are path-dependent. Indeed, institution designers are often repairing problems that emerged during the failure and breakdown of earlier institutions. Therefore, they try to anticipate the effects of institutions, to block bad paths and facilitate good paths, and to create institutional mechanisms to deal with what they cannot anticipate.

Several of the Rational Design dimensions of institutionalization—flexibility, control, and centralization—address how institutions can be designed to adapt to changing circumstances, including those that could never be fully anticipated in the original design. The flexibility conjectures, in particular, describe how institutions accommodate the unforeseen changes represented by uncertainty. Here, individual actors are empowered to adapt their behaviors to new circumstances but must do so in ways that respect and maintain the stability of the institutional arrangement. The centralization and control dimensions suggest ways where the institution itself is empowered to adapt its rules to deal with new and unanticipated circumstances. Thus a secretariat may be authorized to implement the goals of an organization in a particular operational activity, or a dispute panel may interpret a rule over which members disagree. For important issues, the members sometimes act directly by voting on some new activity or rule.

Thus the Rational Design understanding of how institutions develop departs somewhat from the view held by Robert Axelrod and Robert Keohane, who recognize but de-emphasize the prospects for rational institutional design. They characterize the construction of international regimes as "groping," whereby states rely on "experimental, trial-and-error efforts to improve the current situation based on recent experience."[43] We agree that evolutionary selection (or groping) is an important aspect of dynamic change, but we emphasize self-conscious rational design as central to the creation, maintenance, and further development of international institutions. Furthermore, we believe that Wendt's emphasis on the opposition of evolutionary models to rational design obscures important complementarities in practice. Rational design can take advantage of evolutionary good fortune by consciously reinforcing and imitating successful designs. More proactively, policymakers sometimes experiment with alternative designs and select "winners" while letting "losers" go. Although parallel experiments are usually limited to small projects (by operational organizations, for instance, like the World Bank or the Food and Agriculture Organization), many international institutions have been shaped by

43. Axelrod and Keohane 1986, 251.

"experiments" implicit in their past operations. Rational designers learn by experience. They stick with arrangements that work and modify ones that do not. This is not to deny the relevance of unplanned events, but good institutional design encompasses centralization, flexibility, and other features designed to address and take advantage of the unanticipated.[44]

Rational design is even relevant to cases where institutional formation has been "spontaneous." Consider sovereignty and the states-system itself. Although these relationships emerged spontaneously around interstate practices, states have subsequently formalized and ratified the implicit rules that emerged. The same has happened with customary law more generally. It has been codified to make it more precise and to ensure that the expectations it entails are widely understood. In response to changes such as those in communications, transportation, the environment, and military technology, formal negotiations have been used to modify and update international law, including the rules of sovereignty.

Furthermore, institutions that have any centralized capacity will themselves become adaptive agents. Their institutional goals are not identical to those of states, but they must attend to states' interests to further their own goals. International organizations like the World Bank and the IMF that need continuing support and funding from states must respond to their needs and increasingly to the needs of nongovernmental organizations that are able to pressure states.

Finally, we disagree with Wendt's assertion that rational design cannot address what he calls "invisible hand" functionalism, where beneficial outcomes are explained by structural features of a system. Wendt's view seems to be rooted in an insistence that rational design must directly govern the behavior of actors all the way down to the microlevel. This is not necessary: Rational design often concentrates on creating macro structures that operate primarily by channeling decentralized behavior. Many of the best designs lay out only broad parameters and allow the details to be filled in by the individual actors. The familiar example, of course, is the market. Markets often emerge spontaneously, but when they do not, they may be promoted by institutional design in terms of the creation of property rights and arrangements to facilitate exchange. Institutional design can also be used to break down barriers to markets that may have emerged for other reasons or to serve certain vested interests. The development of the GATT/WTO can be seen partly in this light, whereas the effort to introduce tradable pollution-emission rights into the Kyoto Protocol is an explicit effort at "invisible hand" design. Again, rather than try to separate invisible-hand design from rational-design explanations, we believe it is more important to understand their interaction.

In summary, although equilibrium analysis focuses on prevailing circumstances, it provides important insights into change as well. Pressure for change can be understood in terms of the shortcomings of existing equilibria; the resolution of

44. For an overview of the complementarities and tensions between rational choice and evolutionary approaches, see Kahler 1999.

these pressures may lie in movement toward a new equilibrium. Of course, equilibrium analysis is hardly sufficient for a complete understanding of institutional transformation and change. But we cannot fully understand them without it.

Conclusion

Our research program, though ambitious, is far from comprehensive. Wendt identifies several significant gaps in our analysis. We close by discussing two that we agree with—the need to evaluate the effectiveness of institutions and the value of normative analysis—and one that we think is severely misplaced—the notion that we have been looking in the rearview mirror.

Institutional Effectiveness

Wendt correctly observes that even if institutions are designed according to rational-design principles, in practice these principles might themselves be flawed. We agree that evaluating institutional designs for their effectiveness is an important task that should be closely related to ongoing work on institutional design. This topic was controversial at the conferences leading up to this volume. We decided to bracket considerations of effectiveness to concentrate on the design variables. Nevertheless, our ability to explain regime design based on the assumption of effectiveness does provide indirect support for the assumption itself.

The case studies report at least indirectly on effectiveness. For example, Mattli shows that the growth of international commercial arbitration has occurred precisely because it provides a more effective way for private parties to resolve certain types of disputes. Mitchell and Keilbach explain how the Rhine chloride convention was effective in obtaining French and German support for Dutch efforts to limit pollution. Oatley shows how the specific institutional arrangements of the European Payments Union were tied to its effectiveness in solving distributive problems among its member states. Indeed, the empirical discussions of design features in practice must ultimately entail some evaluation of the effectiveness of the institutional arrangement. Nevertheless, we agree that more detailed study of effectiveness—including an analysis of whether states change designs when effectiveness is inadequate—would be a valuable adjunct to our analysis.

The study of effectiveness will be especially interesting in situations where alternative designs are available to address a problem. For example, an enforcement problem might be addressed by membership restrictions, issue scope, centralization, or some blend of these institutional features. But theory alone cannot tell us which combinations will be chosen. Thus the empirical study of effectiveness goes directly to the more complex questions of design interactions—both as substitutes and

complements—that our simple bivariate conjectures do not capture.[45] Moreover, effectiveness raises the issue of whether there are design "traps" whereby a sub-optimal institutional design equilibrium is sustained even though an alternative and superior design is available. Such research requires not just further intensive comparative analysis of institutions but also the study of institutional development over time as actors rearrange designs to better address the problems they face.

Finally, Wendt's implication that rational design is not relevant when we do not understand either the "objective" world or our "true interests" (his terminology) is unwarranted. As we argue in the introduction, one of the key purposes of creating more centralized institutions is to reduce uncertainty about the "state of the world" (conjecture C2). International organizations, for example, sponsor international research and information exchange to promote learning about problems. International conferences and soft law processes are often aimed at promoting more general understandings not only of problems but of the common interests of the participants.[46] Thus rational design can address the issues of uncertainty that Wendt and we agree are so central to international institutionalization.

Normative Considerations

With respect to Wendt's emphasis on the need to broaden our analysis to address normative considerations, he is correct that any rational-design approach implicitly includes a normative element—through the goals of the designers, through the interests to which they must attend, and perhaps through its very orientation toward action. All merit serious consideration. Rational design focuses on incentives (though not necessarily material ones) for solving problems such as collective action. But it is essentially silent on the sources of preferences that underlie these incentives, and on the development of collective values that might guide institutional design. This is not a problem of rational design per se but of the general issue of linking positive and normative analysis. Rational design is no more handicapped in this than are other approaches, and, we argue, it brings some advantages to normative analysis.

In several places Wendt suggests that the logic of appropriateness is more closely attuned to normative issues than is our rational logic of consequences. While normatively attractive notions of "right" (or legitimacy, duty, fairness, and so on) may provide the basis for logic of appropriateness, the reverse need not hold. Slaveholders, Gestapo members, and colonial powers all act "appropriately" in terms of their own identities but not in terms of at least some normative criteria. Similarly, rational design can be motivated by normatively attractive goals—fairness, equity, altruism—but it also can be motivated by morally reprehensible

45. For a fuller discussion, see the volume's introduction.

46. Abbott and Snidal 1998 and 2000. At a more technical level, Wendt misconstrues the implications of the "theory of the second best." Second best theory does not recommend that we stop trying to "maximize expected utility." Instead, it provides guidance on certain pitfalls in doing so.

goals. So we agree with Wendt that rational design is valuable and perhaps even necessary for achieving normatively attractive goals, but it is not sufficient.

Wendt overlooks another potentially important link between rational and normative analysis. Efficiency is a core value of rational design and it carries substantial normative weight of its own. Whatever our specific goals, rational design can help us achieve them better. In this limited sense, the positive connotation of "rational" is not accidental.[47] Moreover, if a normative theory of international institutions is to be practical, it must deal with the issues covered by rational design. Normative considerations should guide what we design our institutions to accomplish, but that design must also take into account the incentives of the actors within those institutions.

Rational Design Is Inherently Forward Looking

Finally, we reject Wendt's suggestion that rational design is akin to "driving with the rear view mirror." We would characterize the Rational Design research program as using past experience along the road—augmented and informed by theory—not only to look ahead but also to plan for unforeseen conditions around the next curve and over the next hill.

Wendt argues that "any theory . . . that only explains past choices will be of limited value in making future ones." But since we cannot yet validate our theories against future choices, we have no better alternative. A well-validated theory gives us some confidence in our understanding and justifies projecting the results to novel circumstances. Wendt is correct that under his most extreme form of uncertainty—knowing literally nothing about the future—rational design will not be of much help (beyond proposing institutional arrangements to reduce our ignorance). But neither will anything else. Fortunately, many of the problems we face display enough continuity and similarity to past problems that the insights will transfer. Rational choice provides a good vehicle for this transference.

Moreover, a number of the institutional features we introduce—notably, centralization, control, and flexibility—are fundamentally geared toward managing various types of uncertainty about the future. Although we derive our conjecture about them in terms of uncertainty viewed as risk, these same institutional features would likely be valuable for handling the "deeper" types of uncertainty that Wendt suggests. If we do not know exactly what lies around the next curve, we should choose the driver or the institutional arrangements that have been shown to be most able to adapt and respond to unanticipated circumstances.

To ask that a design be forward looking is not to suggest that its designers be omniscient. To say that the American Founders did not anticipate today's U.S.

47. In fact, one of the exciting areas of research in international relations has been on the role of normative factors in affecting international outcomes through persuasion and direct action. Margaret Keck and Kathryn Sikkink make the important point that normative actors must act strategically and take into account issues of institutional design if they are to achieve their goals. In short, normative actors need good designs. Keck and Sikkink 1998.

Constitution is not to imply that their design was not considerably better than the alternatives (such as the Articles of Confederation) or that their design did not have a profound impact. The brilliance of the U.S. Constitution, and the reason for its wide emulation, is that it solved a number of fundamental problems facing the country in the late eighteenth century. It also proved effective and durable enough that later generations were unwilling to abandon it even when it failed to meet all their needs. It remains an important institution because it contains provisions for interpreting and adjusting it to new conditions. Indeed, one of the key points of rational design is precisely that, because we cannot anticipate the future, we must design institutions capable of managing the problems that arise down the road.[48]

Finally, a significant tension within rational choice is explaining why, if a superior institutional arrangement is possible, actors have not already adjusted to it. Why study rational design except to document what has already been achieved? But this misinterprets a central contribution of rational-choice theory: demonstrating that rational individuals in collective decision-making settings do not automatically achieve their optimal outcomes. The reasons are several—including multiple equilibria, path dependence, and error. But perhaps the most important reason is the existence of a hidden informational constraint: actors do not know which institutional designs will solve particular problems.[49] The learning process in international relations has been uneven precisely because neither state leaders nor scholars fully understand institutional design—witness, for example, the tragic failures of Versailles and the League of Nations and the imperfect achievements of the UN. But compare these with the successes, like the WTO, NATO, and the EU—from which practitioners and scholars can draw lessons for future designs.

The best way to incorporate these empirical lessons is by combining them with a strong theoretical understanding. This has been the imperfect design of this project. We look forward to having this volume in the public discourse so that other scholars can evaluate the results and continue to test whether our initial conclusions hold.

48. For similar reasons, there is no need to presume that institutional designs must be all embracing. The best designs, like property rights in the market, often establish broad parameters within which individual actors determine specific outcomes through their actions.

49. Townsend 1988.