INTERNATIONAL LAW AND INTERNATIONAL RELATIONS

Edited by **Beth A. Simmons**
and **Richard H. Steinberg**

cannot be certain about the outcomes of their choices, they can at least see well-defined (if still probabilistic) relationships between ends and means, so that they can calculate precisely the chances of achieving their goals with different strategies.[23] Choose *A*, and there is a given chance that pay-off *X* will occur; choose *B*, another chance; and so on. This is significant because it means there is always a clear and principled answer to the question, What is the rational thing to do?

*** [Uncertainty] exists when an actor does not know all the possibilities in a situation, cannot assign probabilities to them,[24] or those probabilities do not sum to unity. To distinguish it from the standard view, uncertainty in this heterodox tradition is often qualified with adjectives like "strong," "hard," "genuine," or "structural." *** [Where] there is genuine uncertainty, the clear (if probabilistic) relationship between ends and means breaks down, so that optimal behavior may not be distinguishable from sub-optimal. If optimality is no longer calculable, then what is instrumentally rational is no longer well defined.

This suggests a rival hypothesis about how rational actors should behave. On the orthodox view, actors facing incomplete information should continually adjust their beliefs and strategies in response to changing estimates of the situation. The importance of such updating is reflected in the volume's conjectures about the effects of uncertainty on rational design, namely that institutions should maximize flexibility and individual control. In contrast, Ronald Heiner argues on heterodox grounds that actors facing genuine uncertainty may be better off *not* trying to optimize, because they are not competent to grasp the true problem and so are prone to make mistakes and have regrets.[25] On his view, in other words, in situations of genuine uncertainty expected-utility theory may actually be a poor guide to "rational" behavior. Instead, actors should do just the opposite of what that theory recommends: follow simple, rigid rules and avoid continually updating expected values. Heiner argues further that most people in the real world understand this, since their behavior is much more stable than would be expected if they were constantly optimizing. Under conditions of genuine uncertainty, it is our willingness to *depart* from the optimizing standard that is the "origin of predictable behavior."[26] In the context of institutional design, therefore, the rational

---

[23] Beckert 1996, 819.
[24] Which may presuppose a nonsubjectivist view of probability.
[25] Heiner 1983.
[26] Ibid.

action may be to *minimize* flexibility and control rather than to maximize them.

*** [The] inferences drawn in the empirical articles about how "uncertainty" should play out concretely seem generally persuasive, and so it is not clear that the heterodox view would lead to different conclusions. Yet some interesting questions remain. In particular, one wonders whether the apparent empirical strength of the volume's treatment of uncertainty is related to the fact that five of its eight articles concern the economic issue-area.[27] One might expect this domain to have relatively weak logics of appropriateness, and so actors will have little incentive to bind themselves to inflexible rules over which they lack individual control. *** However, in issue-areas where logics of appropriateness are stronger, like human rights or perhaps the environment, the heterodox view may be a better guide to "rational" design. In the face of (genuine) uncertainty in these domains states may prefer to define rigid criteria of acceptable behavior rather than maintain the conditions for optimizing their individual interests. On this continuum the security issue-area may occupy an interesting middle ground: in some respects a domain of pure rational self-interest where the volume's conjectures should apply, in others one of deep if limited norms, like those embodied in Just War theory and prohibitions on the use of chemical and biological weapons, which seem harder to square with a desire to maintain flexibility and control.[28] In short, the possibility that the meaning of rational behavior under (genuine) uncertainty varies by issue-area seems worth pursuing. ***

* * *

Alternatives to "Design"

In the preceding section I mapped some of the contrast space implied by "rationality" as a determinant of institutional variation. Although there will be some overlap, doing the same for "design" will put the volume in different relief.

* * *

Thinking about rational design as essentially equivalent to rational choice is also useful for mapping contrasts to the design hypothesis.

---

[27] Of the remainder, one (Mitchell and Keilbach) does not address uncertainty much at all, and another (Kydd) does so in a somewhat idiosyncratic way due to the problem being addressed.

[28] On the chemical weapons case, see Price 1995.

Intuitively the idea that designs are choices has three implications: (1) designers exist prior to designs, (2) designs are intended, and (3) designers have some freedom of action. Each points to alternative explanations, some rivals to rational-design theory and some with greater causal depth. I take these up in turn.

No Designer?

Are institutional designers causes or effects of their designs? On one level the answer must clearly be causes. Institutions do not come out of the blue but are designed by people. However, on another level we can also see the reverse logic at work, with designers being constructed by designs. To that extent, perhaps more is going on in institutional design than the rationalist lens captures.[29] Designers could be constructions of designs in two ways, causally and constitutively.

First, institutional designs today may play a causal feedback role in constructing the actors who make designs tomorrow. This could occur on three levels. As Koremenos, Lipson, and Snidal briefly note, one level would be institutional designs that expand the set of members who make up the subsequent designing actor. In their example of the EU, enlargement choices made in the past affected who is making enlargement choices today, and this will affect who makes choices in the future.[30] A second kind of feedback on actors occurs when institutions affect designers' identities and interests. NATO is a good example: Even if its original design reflected the self-interests of its members, over time they arguably have come to identify with the institution and thus see themselves as a collective identity, valuing NATO as an end in itself rather than just as a means to an end.[31] *** And third, institutional designs may affect actors by changing their beliefs about the environment. *** Such feedback effects may not be intended at the moment of initial design, but the longer our time horizon, the more likely they will occur. Over time, designs cause designers as much as designers cause designs.

The rationalist approach can also be turned around in a second, more constitutive way by adopting a "performative" model of agency. On this view, associated especially with post-modernism,[32] there is an important

---

[29] For further discussion of this idea, see Wendt 1999, chap. 7.

[30] Koremenos, Lipson, and Snidal 2001, 778.

[31] See Risse-Kappen 1996; and Williams and Neumann 2000.

[32] See especially Ashley 1988; Campbell 1998; and Weber 1998. For critical discussion, see Laffey 2000.

sense in which actors do not preexist actions, but rather are instantiated as particular kinds of subjects at the moment of certain performances. To the extent that they are not separable, actors cannot be said to *cause* institutional designs, but are instead constituted by them.[33] In international politics the institution of sovereignty provides perhaps the most fundamental example. By acting as the members of sovereign states are expected to act – defending their autonomy, privileging their citizens over foreigners, recognizing the rights of other states to do likewise, and, now, engaging in practices of international institutional design – certain groups of individuals constitute themselves as the corporate actors known as "sovereign states," which have particular powers and rights in international politics. *** Since this process is continuous, state identity is always an ongoing accomplishment, not ontologically given.[34]

* * *

 This ongoing process of constructing modes of subjectivity matters for at least three reasons.

First, it is part of what is "going on" in institutional design, and therefore a complete understanding of the latter must address it. Doing this would enable us to embed the rational explanation within a larger historical process in which institutional designers are themselves at stake in their practices.

Second, institutional design creates and reproduces political *power* – since in making choices designers are constituting themselves and others as subjects with certain rights – and as such studying the construction of designers by designs matters normatively. Designing a POW regime helps legitimate the right of states to make war and thus kill members of other states; designing a trade regime helps legitimate states' right to protect private property even if this conflicts with justice; and so on. "We" might want states to have those powers, but then again we might not; and our preference may depend on who is included, and excluded, in this "We." Constituting states and their members as the bearers of sovereign rights is an intensely political issue, and so bracketing it in favor of an assumption of given state subjectivity de-politicizes the design of international institutions to that extent. Calling attention to the effects of designs on designers is a way to ensure the power of the latter remains accountable rather than being taken for granted.

---

[33] For discussion of this distinction, see Wendt 1998 and 1999, 77–88.
[34] Ashley 1988.

Finally, this issue raises questions about rationality. If part of what institutional designers are doing is choosing future designers, how do we assess the rationality of the choices they make today? The Rational Design framework defines rationality relative to a given conception of Self. This is fine for certain purposes, but what do we do if the Self will change as a result of our choices? Do we factor in the preferences of future, as yet nonexisting, designers, and if so, which ones and at what discount rate?[35] Attending to alternatives to the assumption that designers are given in design choices would push these important questions to the fore.

No Intentionality?

A second assumption implied by the Rational Design framework is that the features of international institutions are chosen intentionally, by a conscious or deliberate process of calculation. At first glance it is hard to see what a plausible alternative to this would be, since human beings are not automatons. As such, there will always be *some* intentionality in the process by which institutions are created. However, this does not mean we can automatically conclude that institutions are intended. In social theory a long and sometimes fierce battle has been waged by proponents of a rival, "evolutionary" explanation of institutions, especially Friedrich Hayek and his intellectual descendants, against the design approach (which ironically they term "constructivism").[36] The intensity of the resistance stems not only from a theoretical disagreement about what explains institutions but also from the perceived political implications of those explanations. Evolutionists argue that in fact it is very difficult to intend institutions, and that failure to recognize this has led to over-confidence and some of the most catastrophic design failures in history, namely communism and fascism.[37] As an alternative to "constructivism" they favor trusting instead to processes of trial-and-error learning and natural selection, which operate like an "invisible hand" behind the backs of rational actors. ***

Proponents of the evolutionary approach do not necessarily deny that people are intentional beings, that we make rational choices, or even that we should tinker with existing institutions. Many would best be described as "rationalists" themselves. Their concern is rather that

[35] For suggestive treatments of these issues, see MacIntosh 1992; and Stewart 1995.
[36] No relation to "constructivism" in IR. For introductions to this debate, see Hayek 1973; Ullmann-Margalit 1978; Prisching 1989; Hodgson 1991; and Vanberg 1994.
[37] See especially Hayek 1973; and Scott 1998.

even though we may be able to modify institutions incrementally to better realize our ends, the limits of human knowledge and cognitive capacity are so profound that we should not think we can intend successful institutions *up front*. Even the most deliberately created institutions, like the U.S. Constitution, have been amended repeatedly since their founding. Each amendment to the Constitution was certainly intended at the time it was adopted, but in what sense is the result of those changes intended, and who was doing the intending? Perhaps the Founders, whose "original intent" has guided the evolution of the Constitution, and who also consciously created a mechanism for amending it. But it would be odd to say that the Founders "designed" today's Constitution, since they could not have anticipated the changes that have been made; in many respects it is clearly an *un*intended consequence of earlier choices. The assumption that institutional designs are intended, therefore, is ambiguous about whether it refers to the discrete changes made at each step of the way, or to the development over time of the overall structure. Intentionality at the local or micro-level is fully compatible with no intentionality at the global or macro-level. ***

Uncertainty is central to the Hayekian argument, and so the Rational Design project's focus on this factor would seem to put it squarely on the evolutionist side of this debate. Yet the introduction and two of the empirical articles make claims that confuse the issue. Specifically, Koremenos, Lipson, and Snidal argue that even institutions that have evolved very incrementally can be explained by the theory of rational design if their rules have periodically been the object of conscious choice.[38] Their example is sovereignty, the features of which today are the result of many changes made intentionally to the original Westphalian rules. Rational-design theory may shed light on some of the micro-level causes of these changes, but do the editors mean to suggest that sovereignty as we know it today was "intended" in 1648, or that all the individual designers of sovereignty since 1648 add up to a single, trans-historical designer? Presumably not, but in that case then the structure of sovereignty today would require an *additional*, nonintentional explanation. Similarly, Mattli argues that the development of international private arbitration can be explained by an evolutionary process whose outcome is equivalent to what would have been achieved by a direct effort at rational design.[39] That may be true, but how

[38] Koremenos, Lipson, and Snidal 2001, 766.
[39] Mattli 2001, 923–24.

is it evidence for rational-design theory? The latter assumes rational actors; evolutionary arguments, in contrast, require no such assumption. The decentralized, unintended process Mattli describes is precisely what evolutionists see as a *rival* to design explanations; it is the structure of the evolutionary process, not the choices at each step of the way, that explains the overall outcome. Finally, in response to the criticism that decision makers may not understand the design problem and as such need to figure things out incrementally, Mitchell and Keilbach suggest that a "trial-and-error process of design, though taking longer, is no less rational or purposive."[40] This again seems to conflate the intentionality of micro-decisions with the intentionality of the macro-result.

Perhaps what these authors are getting at goes back to their functionalism: If, over time, actors make intentional changes to an institution such that the overall result is functional, we can say it was "designed." Yet this seems to introduce a new understanding of "functionalism" from the one underpinning this volume. If micro-intention equals macro-intention, we seem to be saying that subjective rationality equals objective (or "trans-historically subjective") rationality. But that cannot be right. Incremental changes may cause institutions to evolve in an objectively functional way, but that evolution is more a behind-the-backs process than a purposive one, and as such would have to be explained by the *structures* in which intentional agents are embedded, not their intentions themselves.[41] If we continue with functionalist imagery, therefore, it may be useful to distinguish two variants: "intentional" functionalism, where outcomes are explained by the expected results of intentional action, and "invisible hand" functionalism, where beneficial outcomes are explained by structural features of a system. Rational-design theory as currently formulated would not explain the latter.

*   *   *

No Choice?

Finally, "design" seems to imply that designers have the freedom to act otherwise, that their designs are "choices." To be interesting this needs to be more than just an existential freedom. Assuming free will, human beings always have the trivial ability to "just say no," even if this means they will be shot. The claim needs instead to be that actors have *genuine* choices to

---

[40] Mitchell and Keilbach 2001, 906.
[41] For good discussions of these issues, see Ullmann-Margalit 1978; and Jackson and Pettit 1992.

make, especially if we are going to use the aesthetic term *design*, which suggests a creative expression of inner desire, where the designer could have done things differently but chose not to.

Some philosophers have questioned whether rational-choice theory is compatible with genuine choice, arguing that its model of man is mechanical and deterministic, reducing actors to unthinking cogs in the juggernaut of Reason.[42] *** Rather than pursue this argument, however, I will take at face value the assumption that institutional designers make choices, and focus on how they might be prevented from doing so by structural constraints. The potential effects of such constraints are captured by two alternative explanations, path dependency and teleology.

The implications of path dependency for functional theories of institutional design have been explored in detail by Paul Pierson ***.[43] Especially when institutions are created piecemeal rather than *ex nihilo*, would-be designers may face a substantial accumulation of existing norms and practices. Such historical structures facilitate elaboration of existing norms through a logic of "increasing returns,"[44] and inhibit adopting norms that would undo them. *** Whether for consequentialist or normative reasons, therefore, actors may be constrained by existing structures from making ideally rational choices and as such get locked into a path of institutional "design" that effectively takes away their choice in the matter. ***

The path-dependency perspective suggests a second alternative to the assumption of choice: the teleological view that institutional designers are really just working out the details of some "central animating idea."[45] This could be interpreted in two ways. One version is that the evolution of institutional designs is driven in a counter-rational direction by the unfolding logic of foundational normative principles like equality, democracy, or sovereignty. *** If the EU continues its current (if halting) institutional evolution in the direction of a federal as opposed to unitary state, for example, then in retrospect one could argue that its core commitment to the principle of state sovereignty contained within it the seeds of the outcome (a federal state being more compatible with sovereignty than a unitary one). At the moment of each decision in this evolution actors might have the freedom to make choices, but in the end, at the macro-level, the overall result was pre-ordained. This brings us back to the earlier

---

[42] See Wendt 1999, 126, and the references cited there.
[43] Pierson 2000b; see also Pierson 2000a.
[44] Ibid.
[45] This alternative is raised by Robert Goodin. Goodin 1996, 26.

discussions about the relationship between designs and designers, and design versus evolution. If designers are merely implementing the logic of norms, what really is doing the causal work here: agents or structures? ***

However, there is another way to spin a teleological explanation that parallels the volume's functionalist approach, suggesting that the two accounts might be compatible. One could imagine a teleological explanation that took as its central animating idea not substantive principles like sovereignty or democracy, but the principle of instrumental rationality itself.[46] ***

In summary, because the Rational Design project does not engage in a dialogue with alternative explanations, it is difficult to assess fully its own explanation of institutional design. ***

BROADENING THE SCIENCE OF INSTITUTIONAL DESIGN

Up to this point I have taken as given that the question we are trying to answer about international institutional design is the positive social science one: What explains the choice of designs? In the rest of this article I raise two questions that are not asked in this volume – about institutional effectiveness and normative desirability – and as such my discussion turns more purely external. *** [Part] of what makes the issue of institutional design compelling is that it does raise big questions beyond the explanatory one. These form another kind of contrast space, the mapping of which will help put the project further into perspective. ***

Let us assume that we want to contribute to institutional design in the real world to be "policy-relevant." *** What should social scientists do to make our study of this issue as useful as possible? In short, what should count as "knowledge" about institutional design?

To answer this it is useful to step back and ask, what kind of "problem" is institutional design? What do we need our knowledge *for*?[47] There is no single answer, but any satisfactory one should recognize first that making institutions is about what we should do in the future. In contrast, explaining institutions is about what we did in the past. By identifying constraints, explanations of the past may provide some insight into the future, but the connection is not straightforward. Consider the implications if rational-design theory were a perfect, 100 percent true explanation of past institutional designs. In that case it would reveal laws of human

---

[46] Cf. Meyer et al. 1997; and Boli and Thomas 1999.
[47] Cf. Wendt 2001.

behavior with which we can predict institutional choices in the future. That kind of knowledge is great for social scientists, but how does it help institutional designers? They do not need a theory to tell them what they are already going to do. Ironically, rational-design theory seems like it would be more policy-relevant if it were false, since then it could be used normatively to persuade decision makers to be more rational next time. ***

From a practical perspective, in other words, it is not clear what the "problem" is to which rational-design theory is the solution. In fairness, this is not unique to this theory: any theory, rationalist or constructivist, that only explains past choices will be of limited value in making future ones. This stems from a basic assumption of positive social science: that the universe is causally closed and deterministic, and so there must be some set of causes or laws that explains why we had to do what we did. To be sure, the complexity of the social world is such that we can rarely know these laws with certainty, and thus our knowledge will usually be probabilistic rather than deterministic. But this is typically viewed as an epistemological constraint, not an ontological one. I suspect few positive social scientists would say that social life is *inherently* nondeterministic in the way that quantum mechanics suggests micro-physical reality is; * probabilistic laws are simply a function of the limits of our knowledge in a complex world. It is hard to see where human freedom and creativity come into such an ontologically closed picture, except in the "error term." In contrast, the basic premise of real-world design is that the future is open, that we have genuine choices to make, that voluntarism rather than determinism rules the day. This openness means that the question of what will happen tomorrow is to a great extent fundamentally normative rather than positive. ***

In short, there is an irreducible ontological and epistemological gap between explaining institutions and making them, rooted in their different orientations toward time ***.[48] Interestingly, this gap between backward- and forward-looking thinking is implicit in E. H. Carr's characterization of the difference between "realism" and "utopianism."[49] As is well known, Carr criticized pure utopianism for "ignor[ing] what was and what is in contemplation of what should be," and thus as being too voluntaristic and dangerous.[50] However, Carr's critique was ultimately

---

[48] On the difference between prediction and forecasting, which are rooted in explaining, and "making" as ways of thinking about the future, see Huber 1974.

[49] Carr [1939] 1964.

[50] Ibid., 11.

not of utopianism per se, but of utopianism untempered by an appreciation for constraints. In his view pure realism was also problematic because it was deterministic and sterile, unable to do anything more than reconcile us fatalistically to the evils of the world. As a result, "sound political thought and sound political life will be found only where *both* have their place."[51] Which one should be emphasized at a given time depends on historical conditions. While sometimes "realism is the necessary corrective to the exuberance of utopianism, … in other periods utopianism must be invoked to counteract the barrenness of realism."[52] With the Cold War over, the international community can once again contemplate the utopian side of life, and this volume brings welcome rigor to that impulse. Yet the way it has posed its central question seems still caught up in a realist mentality, oriented toward explaining rather than making, determinism rather than voluntarism.

*** The different temporalities of explaining and making mean there will always be a gap between a science of the past and a policy for the future. If we want to drive forward rather than just see where we have been, therefore, we need kinds of knowledge that go beyond the causes of institutional design, and we need two in particular: knowledge about institutional effectiveness and knowledge about values.

Institutional Effectiveness

Functionalism assumes that actors will choose those institutional designs that they believe will most efficiently serve their interests. As such, the criterion for whether or not an institution is a rational choice is subjective (at the level of the group), namely that it helps them solve their perceived collective-action problem. ***

However, institutions are designed to solve problems in the world, and therefore we will also want to know how well they fit or match the reality toward which they are directed. If institutions perform as their designers expected, there is no problem. Functionalism would then correspond to a Dr. Pangloss situation, the best of all possible worlds. But what if designers' expectations turn out later to have missed the mark? What if an institution has unintended negative consequences of sufficient magnitude that had these been known in advance designers would have made different choices? In short, what if design features are not, in fact,

---

[51] Ibid., 10; emphasis added.
[52] Ibid.