CHALLENGES TO RATIONAL CHOICE

Rational choice is one of the major approaches to the postwar study of international relations (IR). It has helped define contemporary theoretical debates about international politics and has advanced our understanding of such topics as the implications of anarchy and the possibility of cooperation. In some eyes, rational choice has been on a mission to establish its hegemony over the field and has failed to appreciate both its own limitations and the value of alternative approaches. Several vigorous critiques of this approach — both internal and external — are now well established, and rational choice might as a result appear to be in retreat.1

This chapter evaluates the challenges raised by those criticisms with an eye toward how rational choice can deal with them.2 I begin with a discussion of the critiques and argue that each contains some significant element of truth. But they are more usefully thought of as stimulants to improving rational choice than reasons for rejecting it. I then briefly investigate this claim in terms of a set of the substantive and theoretical puzzles that rational choice and the IR field more generally are currently trying to deal with. Rational choice offers no simple solutions to these problems, and it is not the only way to address them. But it is a powerful and flexible approach that, if it takes the criticism seriously, will rise to these new challenges.

Internal critiques about the way rational choice conducts its analysis can be divided into two primary categories. The first is that rational choice has developed a fetishism over mathematical technique that leads it to substitute abstract and complicated models for commonsensical theoretical development. At its best, this line of argument continues, rational choice simply reproduces what we “already know” in a more obscure language; at its worst, it uses obfuscation to hide its emptiness. Worse yet, explanations not cast in the language of rational choice, and even arguments highly compatible with rational choice but not couched in its technical garb, have not been appreciated. Technique has falsely triumphed over substance.

While it contains an element of truth, this critique fails to appreciate the value of
technical approaches used properly. More importantly, it conflates a discussion of particular ways to do rational choice with the approach taken as a whole, which need not be technically complicated or mathematically obscure. Proponents should realize that formalization is not the sine qua non of rational choice but only a tool. Applications of, and even deductions within, rational choice can be entirely "verbal." That is often the best way to proceed. But skeptics should appreciate that the power of approach derives in no small part from established results that might not have been obtained except through formalization.

The second internal critique is that rational choice has no strong empirical legs. According to this view, proponents have not tested it adequately and, when they have, have found little support. Instead of remediying this deficiency, they argue, rational choice has retreated to theoretical speculations that are increasingly irrelevant.

Again, there is an element of truth here, since some rational choice has been more heavily oriented toward developing theoretical arguments about international politics than toward evaluating them. Moreover, the versatility and flexibility of rational choice create special difficulties for testing. But this critique is mistaken in two important respects. First, it confuses the (im)possibility of testing rational choice as a general approach with the more reasonable project of testing specific hypotheses or substantive theories based on a rational choice perspective. Second, it seriously underappreciates the extent to which rational choice is driven by empirical considerations and plays a central role in a wide range of empirical work. Nevertheless, the greatest challenge to rational choice is to strengthen its range and depth of empirical application. This is an ongoing project, however, and I document substantial work that addresses this challenge. In the end, the most interesting debates should be over how best to evaluate rational models empirically, since there is no disagreement over the need to do so.

More recently, an important set of external critiques, grouped loosely under the constructivist label, have pointed out that rational choice emphasizes certain problems and sets aside other issues by assumption. This leads some to doubt the value of rational choice contributions altogether, whereas others are sympathetic to the contributions of rational choice but see it as having run its course or being unable to answer big questions. Rational choice is found deficient in explaining who the key actors are, in explaining their interests, explaining the origin of institutions, or in explaining how these change. These deficiencies present challenges that need to be understood and, where possible, addressed. But the deficiencies are also overstated and (in their weaker form) justifiable as savvy methodological moves by which rational choice analysis gains its power.

My goal here is not to engage the constructivist–rationalist debate but to draw on constructivist criticisms as posing important challenges that rational choice can and should address. I argue that IR rational choice analysis cannot resolve all of these challenges, but it can improve itself by taking them seriously and selectively modifying itself in response.

I address these different critiques, first from a methodological direction and then, building on that, from a substantive direction. The next section begins by asking "What is rational choice?" I propose that rational choice is a methodology incorporating general theoretical assumptions but that it is wide open in terms of specific substantive content. Indeed, its association with particular substantive positions – especially ones deriving from its tradition in realism, neorealism and, more recently, neoliberalism in IR – makes it too easy to confuse the limits of these substantive approaches as inherent limits of rational choice. In fact, rational choice is extraordinarily flexible and is compatible with a wide range of substantive approaches. Other reputed limitations are not inherent to rational choice but represent tactical methodological choices to facilitate
WHAT IS RATIONAL CHOICE?

Rational choice is a broad enterprise with permeable boundaries. It is not a substantive theory except at the most general level. Therefore, it is usually viewed as a methodological approach that explains both individual and collective (social) outcomes in terms of individual goal-seeking under constraints. This broad conception needs to be filled in considerably before it can have much specific content, either theoretically or empirically. There is a myriad of possible ways to do so – some of which respond to important critiques of the approach.

Yet any theoretical methodology such as rational choice entails some very general substantive commitments. The focus on goal seeking presumes that explanation should proceed in terms of relevant actors, the goals they seek, and their ability to do so. The approach also requires some specification of constraints – which may be technological, institutional, or arise from interdependencies among actors’ choices. Within and beyond this, rational choice is remarkably open to alternative specifications. Notably, the goals are not restricted to self-regarding or material interests but could include other-regarding and normative or ideational “goals.” Moreover, while the baseline theory is often developed in terms of hyper-rational actors with powerful calculating abilities, the theory is open to incorporating limits to their capacities or constraints on their decision making. Finally, the theory is most often used as a positive theory of how actors behave in practice, but it can also be used as a normative theory to evaluate how actors behave or to indicate how they should behave.

Any application of rational choice that aspires to be a theory “of something” requires more detailed substantive commitments (Snidal, 1985). In IR, the “neo-” tradition has assumed that states are the key actors, that they seek goals such as power or wealth, and that they are relatively effective at pursuing their interests. The warrant for
these assumptions comes not from rational choice, however, but from substantive knowledge of international politics. For example, the social choice literature raises thorny questions about treating aggregates as actors (Arrow, 1951) so that use of the state-as-actor assumption depends on implicit substantive claims. Alternative assumptions that propose other actors (for example, transnational activists or subnational interests), or actors pursuing different goals (for example, moral values or profits) or actors with differential abilities to do so (for example, in terms of more limited calculating ability or lack of information), are equally compatible with rational choice. Thus, rational choice is not at all limited to conceptions of self-interested, materialistic “economic” actors, or to anomic, power-seeking state actors in international affairs. Different substantive specifications can lead to different theories within the broader umbrella of rational choice.

Goal seeking obviously does not cover all aspects of human (or state) behavior in any straightforward way. There are other rich traditions of research in IR based on (for example) psychological or cognitive limits of decision makers (Jervis, 1976; Steinbruner, 1974) or, more recently, explanations that depend on identity or culture (Katzenstein, 1996), or on the role of “appropriateness” as an alternative basis for behavior (Finnemore, 1996; March and Olsen, 1989).6

The elasticity of the rationality concept makes it tempting (and a little too easy) to reduce these alternative conceptions to a form of goal seeking. Treating “appropriateness” as an element of utility function, or bounded rationality as information costs, simply misses the difference between the approaches, which need to be taken more seriously (Sen, 1977). Nevertheless, rational choice is often a central part of the explanation even where different motivations are also at play. Many behavioral findings about the limits of decision making – hysteresis and framing effects, for example – are implicitly defined in terms of deviations from a rational choice baseline. Similarly, while human rights activists or other actors may be driven by different considerations and use different techniques than are typically captured by rational choice, the analysis of their behavior in international politics requires careful attention to their strategically rational behavior (Johnston, 2002; Keck and Sikkink, 1998). Conversely, rational choice analysis can advance by taking the alternatives seriously and seeing what elements it can incorporate. An example is the effort to understand communication (Morrow, 1994a) and rhetoric (Goldsmith and Posner, 1999, 2002b) among states. But rational choice does not have to be a closed system in this process, and the resulting explanations may blend elements of rational choice with alternative approaches.7

One criticism of rational choice is that it takes the identity and interests of actors as outside the analysis – and thereby brackets one of the most interesting aspects of international politics and change. Leaving that issue to later, it is not true on the constraint side, which is determined by both exogenous and endogenous factors. Some constraints derive from available “technology,” but the more interesting ones are political and social. In many equilibria, for example, each individual is “constrained” by others’ choices in the sense that its best choice, and what it can achieve, depends on the choices of others. Institutions themselves are equilibria – sometimes emerging endogenously within a game and sometimes the legacy of interaction in a prior game – that serve as constraints for actors in a game (Calvert, 1995; Snidal, 1996). These institutional constraints also provide the means by which rational choice can move beyond its focus on actors (that is, its “ontological commitment”) to investigate how institutions impose structural constraints on the actors. This is especially important with regard to beliefs which straddle individual goal seeking and collective institutional constraints. Beliefs are properties of individuals, but their impact often comes because they originate in and depend on
intersubjective “common knowledge” among the collective (Morrow, 2002; Wendt, 1999).

Operating as a causal theory, rational choice is often criticized for assuming what is of greatest interest – including the identities of the actors, their interests, and the institutional structures or rules of the game. To be sure, a causal rational choice account needs to be clear about what is exogenous and what is endogenous in order to proceed. It may specify these things according to substantive knowledge of which elements are (relatively) fixed compared to other elements more subject to change. It therefore defines the scope of what is being analyzed, and identifies various factors as endogenous (to be explained) or exogenous (taken as given). The latter category includes “deep” assumptions that the researcher uses implicitly without highlighting or perhaps even realizing (for example, market institutions in traditional neoclassical economics; sovereignty in neorealist arguments). In other cases, a “partial” equilibrium analysis reflects an explicit understanding that certain factors are being held constant which might otherwise affect the analysis. The justification for this bracketing is that the excluded effects are small – or change slowly – compared to the effects of the included variables.

Moreover, rational choice analysis is not inherently causal, as is reflected in the centrality of equilibrium analysis in the theory (Marshall, 1910). An equilibrium is a statement of consistency among specified elements, that there is no pressure on any of the elements to change given the values of the other elements. It is thus an evaluation of a whole state of affairs and claims only that the elements can coexist with one another while stipulating nothing about their sequence or causal relation. In this respect, equilibrium analysis is constitutive rather than causal (cf. Lake and Powell, 1999b: 32; Wendt, 1999).

Causal analysis is induced when substantive assumptions of exogeneity and endogeneity are introduced for tactical methodological reasons to trace the implications of change in one element on another while holding other elements fixed or constant. Comparative statics does this by assuming an exogenous change in one element will result in an endogenous change in one or more other elements. But the choice of what is fixed and what fluctuates is not inherent to rational choice but to the interpretation put on the model.

Rational choice is also associated with some further methodological commitments that are neither logically entailed by it nor necessarily distinguish it from other research approaches. One commitment is to simplification, the notion that good explanations are lean and minimize the assumptions made. This simplicity – and the structure of some of the problems that are analyzed – lends itself to formalization, discussed below. Another reason rational choice stresses simplicity is to constrain its own versatility so that its explanations do not become tautologies. But the price of making simplifying assumptions such as fixed interests and fixed environments is that rational choice sets aside potentially important questions (for example, what determines interests) by assumption. Even if this self-imposition of theoretical blinders is for a good reason, it nevertheless raises the question whether rational choice can relax those assumptions (possibly by tightening others) to broaden its analysis.

A second commitment of rational choice is to generalization. One virtue of abstract concepts and models (for example, prisoners’ dilemma, agency problems, asset specificity) is that they transcend substantive problems. Thus, similar rational choice analyses have been offered of such disparate phenomena as the family, the market, and war. This greatly facilitates the transference of insights and intuitions across fields, for example, from the study of American politics to IR (Martin, 2000; Milner, 1998). It has also been an important factor in the spread of IR theory into other areas such as international law (Abbott, 1989; Burley, 1993; Setear, 1999). But this commitment to generalization is not simplistic in the sense of requiring that human action can be reduced to fully generalizable
“laws,” as is often implied by critiques. Instead, it is in the spirit of showing how a very broad framework can encompass compatible analyses that explain many different situations. Indeed, the development of rational choice in its game-theoretic mode has involved a proliferation of increasingly specific and therefore different structures or contexts (that is, game specifications), each of which may exhibit substantial further variation as its parameters vary. Generality obtains at the level of transferable insights rather than mechanical rules or scientific laws.

An immediate implication is that rational choice is not a singular approach but rather a large family of approaches. These are highly complementary at the level of basic methodology, but they can be completely inconsistent in terms of their substantive arguments. Thus, many of the debates within rational choice – and certainly most of those within IR – are actually debates driven by substantive assumptions that originate outside the methodological framework.8 Rational choice can play a useful role in clarifying and even adjudicating these debates because it provides a common conceptual framework for specifying the problem and a machinery for checking consistency and implications of arguments. Gaining the full advantage of these capabilities explains the importance both of formalization of arguments and of developing empirical connections.

**Rational choice and formalization**

Rational choice has an elective affinity for formalization, by which I mean the use of mathematical models to represent theoretical arguments and simplified versions of the real world. Formalization is by no means a necessary feature of rational choice. Many important applications of rational choice can be properly described as “soft,” meaning that they are not closely tied to formal models or arguments. Indeed, one indicator of the success of formal rational choice is the extent to which key analytic arguments and conclusions, such as those emerging from theories of collective action, cooperation theory, principal-agent models, and signaling models, have become part of our verbal vocabulary and common understanding of international politics. This is not to say that mathematics is not central to rational choice – far from it – since many important developments have been generated or significantly improved by formalization. Moreover, the persuasiveness of soft rational choice often rests on the “hard” formalization that stands behind it.9

Theoretical rational choice need not rest on mathematical models. Early rational choice political economists, including Adam Smith and David Hume, did not invoke mathematics. Much of the contemporary economics that has been most influential on IR – notable examples include the work of Ronald Coase (1960), Douglass North (1990), and Oliver Williamson (1975, 1985) – is not mathematical to any significant extent.

**METHODOLOGICAL DEBATES AND CHALLENGES FOR RATIONAL CHOICE**

This section examines “internal critiques,” which are largely methodological in terms of debating “how to do rational choice” rather than the advantages or disadvantages of the approach taken as a whole. I focus on the relation of rational choice to formal mathematical models and to empirical testing. My argument is that formalization is highly complementary to “softer” (that is, nontechnical) rational choice and that neither can be successful without close attention to the other. The complaint that rational choice has paid insufficient attention to empirical testing holds some merit but nevertheless greatly undervalues the substantial empirical work of various types that is based on rational choice approaches. More importantly, attention to empirical issues has grown considerably among rational choice researchers, and a new concern for evaluating theoretical arguments is apparent.
Moreover, traditional IR contained significant though informal elements of rational choice thinking long before any move toward formalization (Morgenthau, [1948]1978: 4–10). Contemporary rational choice theorizing in IR often draws on formal results but is not formal itself (e.g., Keohane, 1984; Koremenos et al., 2001a; Lake, 1999; Oye, 1986).

Conversely, mathematical models need not involve rational choice. Recall Lewis Frye Richardson’s (1960) arms race model using coupled differential equations to represent a dynamic action-reaction process where each nation-state acquires weapons in response to the others’ level of armaments. Richardson describes it as a model of “what men would do if they did not stop to think.” Even game-theoretic models have no necessary connection to rationality. Evolutionary game models are of this sort and their relation to rational models is an area of ongoing research (Kahler, 1999). Barry O’Neill (1999) uses game theory to underpin an explicitly nonrational choice framework – indeed he allies his work closely with constructivism – and to explain a wide range of international phenomena through symbolism.

There is also no singular way to model rational choice. Noncooperative game-theoretic models are the predominant approach, and they largely subsume traditional microeconomic and decision-making approaches. But choice within this broad family of models entails increasingly detailed substantive commitments. In deciding on a specific model – for example, between a strategic game model versus a nonstrategic decision-making model, or between an extensive (tree) form game versus a simultaneous (matrix) form game, or between complete versus incomplete information games – we are making substantive judgments as to what are the most important features of the problem. The tendency to assume that one class of models (often more complex ones or newer ones) is inherently superior is generally misguided. Proper model choice, including the level of formalization, depends on the substantive problem being analyzed.

Simple representations have been as important as technically complex ones in promoting rational choice analysis. The obvious example is the prisoners’ dilemma (PD) game, which is now so much part of our common theoretical vocabulary that Stephen Walt (1999: 9) seeks to exclude it as an example of formalization in his critique of such approaches even though it fits his definition of a formal theory. Like other simple games, including “assurance,” “coordination,” and “chicken,” PD carries an impressive analytical load that has considerably clarified and advanced our thinking about international politics. The fact that it has become a standard metaphor for international anarchy is an indicator of its power as soft theory; the fact that its clarity has made its limitations apparent and stimulated a large family of extensions and refinements shows the power of its formal representation.

Many soft rational choice arguments that have become standard in the literature are ultimately based on mathematical derivations. Although Walt treats Olson’s (1965) analysis of collective action as not placing “much emphasis on mathematical rigor,” that claim seriously misunderstands Olson’s enormous debt to Paul Samuelson’s (1954) very rigorous and quite technical analysis of public goods. The economics field was befuddled by public goods until Samuelson’s formalization provided a clear conceptual basis to move forward. Olson’s great contribution was to connect the formal analytics to wide-ranging political examples and to extend the analysis through soft theory (for example, selective incentives). It provides an excellent example of how soft rational choice can build off formalized work to overturn well-entrenched conventional wisdom and create a common framework for substantial advances.

Similarly, the result that cooperation can be supported by strategies of reciprocity is referred to as the “folk theorem” because a simple version had been widely known (at least among game theorists) for a long time. The understanding of the effectiveness of
reciprocity also predates its mathematization, while explicit “tit-for-tat” policies predate game theory by 400 years, and informal “eye-for-an-eye” behavior goes back much further. Yet the folk theorem result that seems so obvious in the light of game theory was a novel claim for IR theory when expressed in terms of the possibility of “cooperation under anarchy” (Oye, 1986), just as tit-for-tat’s victory surprised most game theorists who participated in Axelrod’s (1984) tournament. Formalization of the cooperation argument was essential for establishing the conditions under which cooperation was possible and for overcoming substantial intellectual resistance on that point. By identifying the precise preconditions for cooperation, formalization also opened up avenues of analysis regarding the limitations of those arguments (for example, poor information, large numbers, distributive issues) as well as ways to overcome those limitations (for example, different strategies, issue-linkage, institutional roles).

While proponents make much of how rational choice formalization has produced results that are “counterintuitive,” critics respond that most of these results are “obvious.” The debate is pointless for two reasons. First, most good results cease being counterintuitive once they are properly understood. Collective action and PD are good examples where what was once surprising is now conventional wisdom (Barry and Hardin, 1982). Second, and more important, counterintuitiveness is vastly overrated as a criterion. In a field that has no shortage of unsubstantiated “surprising” claims, formalization performs a more important service by helping us work out the underlying logic and clarifying the grounds for the different claims. A prime example has been the improved understanding of the role of information in the deterrence (Powell, 1990) and assurance problems (Kydd, 2000).

A related contribution is that formalization provides a systematic framework for exploring detailed implications and extensions of arguments. While this happens within any one model, it also occurs across analyses.

A good example is the set of interconnected advances that follow from the initial formalization of how cooperation is possible among states. For example, Downs and Rocke (1995) show how domestic uncertainty will impede cooperation and the impact this may have on preferred institutional arrangements. James Fearon (1998) shows that when bargaining differences among alternative cooperative outcomes are introduced, the “shadow of the future” that enforces decentralized cooperation among states also creates incentives that may impede the attainment of cooperation. Barbara Koremenos (2001) demonstrates that introducing flexibility to cooperative agreements provides an institutional means of overcoming such problems in many circumstances. Although the specific models vary, their close relationship and clarity allow their conclusions to be readily compared and integrated.

A different complaint against formalization is that with a little cleverness one can derive virtually any conclusion (Stein, 1999). Unfortunately, the complaint is not generally valid, as is demonstrated, for example, by the difficulty of finding a realistic mechanism for optimal public goods provision and by the difficulty of providing a satisfactory rationalist account of war (Fearon, 1995; Garzke, 1999). Even if a model can be trumped up, however, the whole point of formalization is to insist that its assumptions and logic be explicit so that any “tricks” can be discovered. It is true that tricks can be buried in the mathematics, but they can also be unearthed and more readily revealed as tricks than can tricks in comparably complicated verbal arguments. The explicitness, precision, and clarity of formal analysis lend themselves to this goal.

Usually, of course, it is not a matter of tricks but of different assumptions. Here, formal analysis provides an excellent way to clarify the terms of a substantive debate. An example is the “relative gains” debate. Grieco (1988) argues that liberal theories misspecify the cooperation problem as an absolute gains problem (that is, how much each received),
whereas states also care about relative gains (that is, who gets more). He shows that cooperation is much more difficult under relative gains, although Snidal (1991) demonstrates that even under Grieco’s assumption, the inhibiting effect on cooperation is significantly limited in multilateral settings. Powell (1991) challenges the need to use the relative gains assumption by showing that the relative gains concerns can be induced as a by-product of strictly absolute gains concerns for survival.

This raises an interesting question about the aesthetics of models and theories. Should we prefer Powell’s more parsimonious explanation because it assumes only absolute gains and derives relative gains concerns within the model? This does not by itself invalidate Grieco’s assertion that states seek relative gains for primordial reasons. Indeed, one might equally be able to derive absolute gains preferences as a by-product of relative gains concerns (that is, I want more in order that I have more than you). This difference can be resolved only by finding a way to investigate the assumption itself or by generating divergent predictions from the two approaches. Either way, the formalization of the argument and the logical derivations sharpen the issues at stake (Grieco et al., 1993).12

In addition to the complaint that rational choice can predict anything, another complaint is that rational choice predicts everything and therefore predicts nothing. It is not true of all models, of course, since some give quite precise predictions. Therefore, it is usually expressed in terms of the multiple equilibria produced by the folk theorem result that (under certain conditions) any point of cooperation that makes everyone better off can be attained. But should game theory be blamed for creating this multiplicity, or credited for revealing it and raising new questions about how the resulting indeterminacy is resolved? Regardless, there is a need to narrow the prediction, a problem which the formal theory has had only partial success with.13 Any solution almost certainly requires bringing additional substantive considerations into the theory – perhaps through psychological theories of decision making, or social norms, or historical analyses of path dependence.

Thus, we can endorse and extend to IR more generally Walt’s approving quotation of Schelling’s warning that we should not “treat the subject of strategy as though it were, or should be, solely a branch of mathematics.” But this caveat must be understood in the context of Schelling’s general concern that “the promise of game theory is so far unfulfilled” with respect to his goal of improving “the retarded science of international strategy.” Schelling’s point is both that mathematics isn’t sufficient for developing good theory and that the verbal theory of military strategy that developed without formalization is moribund. His work shows how formal analysis can provide an invaluable basis upon which to build a largely verbal theory. Schelling’s brilliance, of course, is that he is able to keep the mathematics in the background and in the appendices, while connecting it to a rich set of anecdotal examples and broad theorizing.

The challenge for rational choice is to find an appropriate combination of hard and soft approaches. Because the formal theory is now so much more developed, the contemporary mixture does and should include some highly technical pieces. At the same time, the real success of these models depends on several things. One is that the model itself not be so complex that it ceases to be a reasonable representation of what actors are actually capable of doing.14 Another is that the underlying logic be expressed in softer terms, even if that means less precisely, in order to exploit the complementarities between hard and soft approaches. Formal results can provide a new impetus for informal speculation, now informed by a clearer baseline and better able to proceed on problems that defy hard formalization. Finally, an important criterion for gauging the success of a model is whether it can be connected to empirical evidence, to which we turn next.
Empirical evaluations of rational choice arguments

Rational choice has also been criticized as insufficiently attuned to empirical matters, especially testing. This is an important and partly valid complaint. Enthusiasm for the theory’s deductive power sometimes displaces attention to empirical realism and testing. This problem is aggravated by the difficulty or impossibility of observing some key variables such as preferences and beliefs, and by the flexibility of the theory, such that it can be adapted to fit the data in an ad hoc manner. However, the partial validity of the complaint should not be misconstrued as indicating an inherent fault of the approach as much as a failure of application. Below, I discuss some of the significant and growing body of empirical rational choice IR research, both quantitative and qualitative, that illustrates the empirical usefulness of the perspective.

Rational choice as a whole cannot be tested; only specific hypotheses and substantive theories within rational choice can be tested. Rational choice operates at such a general level, and covers such a range of models, that it makes no more sense to think of testing it than it would to test “statistics.” Furthermore, at the most abstract level, the conclusions of rational choice are “true” by logical derivation. They are made empirically relevant – and falsifiable – only by mixing them with substantive assumptions (for example, goals of actors, “rules of the game”). Thus, what can be tested are applications of particular rational choice models to specific substantive problems. If the finding is negative, we reject the application of that model and the associated substantive assumptions and theory to that issue – not the model itself, and certainly not rational choice as a whole. This is the same as (say) when a Poisson model of war initiation fails and we reject the attached substantive theory of war initiation, not the Poisson model or its potential application to other areas of international politics. Of course, if a Poisson, signaling, or whatever type of model fails across a wide variety of circumstances and succeeds on few, we should conclude that it isn’t much help for understanding IR. Only in this very limited sense can we “test” rational choice as a whole.

For this reason, much of the antagonism between rational choice and its critics is again misplaced. Rational choice should not claim to explain everything and should not be held to that standard. The appropriate question is whether it can explain a fairly wide range of phenomena. Although the standard IR approach is to line up one theory against another so that the success of one is tied to the failure of another, that is not a necessary condition for understanding international politics. It is perfectly plausible, bordering on the obvious, that actors are motivated by both the “logic of consequences” and “the logic of appropriateness,” for example, and our empirical task is to sort out under what conditions each logic operates – including the recognition that they operate together in some circumstances (March and Olsen, 1989). It is equally apparent that actors sometimes deviate significantly from the strict assumptions of rational choice so that conclusions need to be tempered by consideration of whether the simplification is adequate for the problem. Conversely, many of these deviations can only be understood in terms of a baseline of rationality. For example, James Fearon (1995: 409) points out that a better rational explanation of war may increase our estimate of the importance of “irrational” factors. So the complementarities of explanations may be more important than their differences – especially if the goal is understanding substantive problems rather than scoring debating points. In brief, rational choice should claim to explain, and be expected to explain, only some important aspects of international life, not all.

A generic difficulty in empirical testing is that when an argument or model fails, proponents will respecify it before abandoning it. This is simultaneously a virtue and a defect, which though not unique to rational choice is
magnified by its versatility. It is a virtue because the “thinness” of any general theory means that its conjectures are unlikely to perfectly match any case or set of cases. Reformulation allows us to capture nuances and details of the case that we did not theoretically anticipate, or to relax simplifying but unrealistic assumptions (for example, replacing the unitary state assumption with a specification of domestic interests). It allows us to revise rather than reject the model. In this sense, it is not unlike the use of different models in statistical work, or different specifications within any particular model. The corresponding defect is that revision undercut testing, and heavy-handed revision merges into curve fitting. This even erodes the logic of discovery if patching up a model to fit a case displaces the effort to identify a very small number of unexpected revisions that improve our understanding. In the extreme case, we end up “rationalizing” the model rather than testing it.16

But while “rationalization” is certainly a concern, rational choice offers safeguards that limit the danger posed. First, formalization restricts curve fitting. As noted, it is difficult to generate a reasonable model that produces a particular result, especially since formalization makes assumptions and arguments explicit and subject to scrutiny. Verbal theory (“soft rational choice”) provides more latitude in this respect, but even here the theory imposes significant limitations on acceptable explanations. Arguments drummed up to rescue the empirical failings of a model are more likely to appear ad hoc when compared to the systematic theory. Second, the emphasis on parsimony works against curve fitting in both formal and verbal theory. (To expand the analogy, it is more difficult to “curve-fit” with a straight line than with a higher-order polynomial, or with one or two independent variables rather than with many variables.) Third, rational choice’s aspirations to generalization limit its ability to modify models to fit particular cases. Finally, and most important, rational choice has adopted methodological strategies to limit the possibility of rationalization. The much-maligned assumption of fixed preferences, for example, arises not from a metaphysical belief that preferences are fixed but from methodological considerations of enabling empirical disconfirmation through limiting the possibilities of “curve fitting” via imputed preferences.17

Thus, rational choice is caught between the “rock” of empirical criticisms and the “hard place” of theoretical (constructivist) criticisms of its conception of international politics. Fixed preferences increase the potential for falsifying a rational model but make it difficult to accommodate change in the character of actors and of the system. However, this difficult trade-off is not unique to rational choice. It emerges from the more general problem of reconciling expansive theory with empirical testability. In other words, the “revealed preference” problem, whereby rational choice explanations tend to tautology if preferences are induced from behavior, is no more or less severe than the “revealed norm” problem of constructivism when social “norms” are induced from observation of social practice.

Although the empirical application of rational choice has lagged behind its theoretical development, a significant and growing amount of attention has been devoted to testing rational-choice-based theories using both quantitative and qualitative methods. On the quantitative side, an excellent example is the emphasis on empirical analysis in the “expected utility” theory of war pioneered by Bruce Bueno de Mesquita (1981). In addition to strengthening the attention to theoretical issues over earlier correlational investigations, this line of analysis has paid careful attention to the difficult measurement problems inherent in expected utility arguments and, especially as it has developed into a fully strategic model, to the difficult problems of estimation associated with equilibrium predictions (Signorino, 1999; Smith, 1999). The implications of the theory are testable, and it has received considerable empirical support (Bennett and Stam, 2000;
Bueno de Mesquita and Lalman, 1992), but also much criticism. The point here is not whether the theory is right or wrong but that it demonstrates that quantitative testing is an achievable goal for rational choice arguments.¹⁸

A second cumulative empirical agenda guided by rational choice has been the analysis of sanctions. The basic model is of one actor (the “sender”) imposing a demand on another actor (the “target”) and threatening to impose a sanction unless the demand is met. In addition to a progression of theoretical efforts to develop improved models of sanctioning (Drezner, 1999; Kirschner, 1997; Martin, 1992; Smith, 1995; Tsebelis, 1990), there has been extensive empirical work in both the qualitative and quantitative traditions to evaluate the theoretical claims (Dashti-Gibson et al., 1997; Drezner, 1999, 2000; Drury, 1998; Pape, 1997; Shambaugh, 1999). Debate continues on the efficacy of sanctions, on the selection of evidence, and on exact theoretical specifications, but there is no debate that this is a serious effort to join theory and data.

On the more qualitative side, rational choice has implicitly underpinned many important arguments which have been evaluated through historical and comparative case studies. One of the difficulties here, as noted earlier, is that rational choice arguments are so pervasive that it is sometimes easier to distinguish analyses that are predominantly not rational choice (for example, organization theory, normative, psychological) rather than ones based on rational choice. Nevertheless, a list of empirical analyses that are closely tied to rational choice theory would include such important books as those by Vinod Aggarwal (1996) on international debt, Jeffry Frieden (1991) on debt and development, Robert Gilpin (1981) on war and change, Hein Goemans (2000) on war termination, Joanne Gowa (1994) on alliances and trade, Joseph Grieco (1990) on the Tokyo Round, Lloyd Gruber (2000) on NAFTA, David Lake (1999) on US foreign policy, Lisa Martin (2000) on making credible commitments, Walter Mattli (1999) on regional integration, Ronald Mitchell (1994) on environmental compliance, Helen Milner (1997) on domestic politics and IR, Andrew Moravcsik (1998) on European integration, Kenneth Oye (1992) on economic discrimination, Beth Simmons (1994) on interwar monetary politics, Jack Snyder (1991) on imperial overextension, Daniel Verdier (1994) on trade policy, and even Stephen Walt (1987) on alliance formation.¹⁹ This is necessarily a partial list of books, and it ignores the much greater number of relevant articles.²⁰

But it certainly illustrates an impressive quantity and range of empirical applications significantly guided by rational choice. It also illustrates a wide variety of styles in the empirical application of rational choice. Some use explicit models to guide their analysis, others use verbal models as heuristics to guide their verbal argument, and still others simply focus on goal-seeking behavior as their fundamental explanatory factor.

Qualitative testing of models based in whole or in part on rational choice has the same possibilities and drawbacks as similar tests of other theories. However, there are two routes (neither exclusive to rational choice) by which this capacity can be enhanced. First, the generality and systematic character of rational choice makes it particularly well suited for evaluation through connected case studies. The potential for development in this regard is seen in comparing the early Cooperation Under Anarchy project (Oye, 1986) to the recent Rational Design project (Koremenos et al., 2001a). The Oye volume shows the value of using very general rational choice insights (for example, “Iteration promotes cooperation in PD”) to guide understanding across a wide range of empirical cases. The RD project builds on this by increasing the specificity of independent and dependent variables and developing a series of precise conjectures based directly on formal rational choice results (for example, “Institutional flexibility will increase with uncertainty”). These institutional design conjectures can be clearly
evaluated in specific cases and are sometimes found wanting. However, the overall success of the conjectures demonstrates the value of rational choice for understanding international institutional design.

Second, Robert Bates et al. (1998) have coined the term “analytic narratives” to describe a more systematic use of rational choice as a qualitative empirical research tool for individual case studies. Their method uses a rational choice framework to guide the description of the case in terms of actors, preferences, choices, environmental constraints, strategic interdependencies, etc., and then invokes stronger theoretical results to assess the underlying mechanisms that lead to equilibrium outcomes. Its distinctiveness rests on a self-conscious effort to connect explicit and detailed rational choice models to historical events. In doing so, analytic narratives engage the tension between the formal logic and parsimony of the rational choice framework on the one hand, and an effort to capture specific contexts and analysis of developments over time characteristic of the narrative on the other hand. This combination has attractive advantages, including opening up the rational choice framework by pushing for a deeper and tighter analysis of the underlying mechanisms of change. The analytic side encourages generalization while the narrative side encourages closer contextual specification. While the approach can be seen as a form of rational choice imperialism, the authors argue it can more fruitfully be understood as an effort to find a “middle ground.”

The rhetoric of “analytic narratives” is exaggerated in a number of respects. First, the approach is not as novel as the name which, as Bates et al. (2000) point out, effectively describes what a number of rational choice and other case specialists have already been doing without the label. In particular, the well-established approach of “process tracing” (George and McKeown, 1985) is similar in spirit and can be applied using a variety of theoretical perspectives (for example, analytical Marxism, prospect theory).

Nevertheless, Jon Elster (2000: 694), in an otherwise unyielding critique of analytical narratives, argues that rational choice has a special advantage because it “is the only theory in the social sciences capable of yielding sharp deductions and predictions.” Of course, this requires that the predictions are sharp — a potential problem if multiple equilibria are pervasive — and that ex post choice among ex ante predictions doesn’t dull the sharpness of the latter.

This richness of predictions poses a challenge for the use of analytic narratives. Proponents present it as a method not only of interpretation and theoretical discovery but of “testing” — although these are rarely so neatly separated in practice — whereas critics may see it as little more than a more formalized approach to curve fitting (Dessler, 2000; Green and Shapiro, 1994). As with any case-oriented approach, this limitation is attenuated (but not eliminated) to the extent that the narrative explains a rich variety of facts within the case and to the extent the same theory fits across cases (King et al., 1994; Van Evera, 1997). Analytical narratives are stronger tools than process tracing or other qualitative approaches to the extent that rationalist theory restricts the range of predictions, and enable skepticism, if not strict falsification, of the application. By this reasonable standard of comparison, analytic narratives hold substantial potential.

Elster’s severe critique of analytic narratives implicitly applies to any systematic explanation that would meet (say) the standards of Green and Shapiro (1994). It nevertheless contains a series of points that should promote modesty in the development of analytic narratives in IR. First, actors should not be endowed with extraordinary powers of computation (for example, able to anticipate complicated strategic contingencies over long periods, or to deal rationally with any type of uncertainty). Overly sophisticated models that require equally sophisticated actors are not plausible representations of the international life, and these complications make testing that much harder. Second, Elster
is suspicious of aggregate actors such as the states, international organizations, or non-governmental organizations (NGOs) that populate many IR arguments. While his preferred solution of disaggregating down to individuals is generally impractical – as well as denying the importance of collective properties that cannot be captured in the individual units – his alternative solution that the “aggregate as actor” assumption be justified is wholly appropriate. This also provides insight into state “preferences,” a point I expand on below. Finally, Elster makes the important point that rational choice models not only must recognize that rational motives need not be material but, equally importantly, that they coexist with other motivations whose impact may not be understood through rational models. Thus, strategic actors can be moral, and moral actors can be strategic. This delimits the scope of rational approaches, but it doesn’t necessarily undermine their power. Proper scope delimitation should make rational choice more effective in the range it applies (Abbott and Snidal, 2002).

Whether in response to its critics or simply as a natural continuation of its research program, rational choice is responding to its weaknesses as an empirically applied theory. The resulting growth in empirical work will have several salutary effects. Closer attention to empirical issues promotes explanations that are “thicker” in specifying key features and context of issues and in identifying underlying mechanisms. While empirical successes in explanation and systematic testing will strengthen the grounds for accepting these explanations, empirical failures will identify aspects of international politics that rational choice does not explain well. Proponents will treat failures as anomalies that need to be accommodated. This will occur first through retrofitting and reformulating the theory – a form of curve fitting which is inherently unsatisfactory from a testing perspective but may be illuminating both of the individual case and of more general mechanisms. The ultimate empirical test, of course, is whether these reformulated models have a more general applicability to other empirical cases.

**SUBSTANTIVE CHALLENGES FOR RATIONAL CHOICE**

The ultimate challenge for rational choice is not whether it has been too formal or has focused insufficiently on empirical matters in the past, but how well it can handle emerging issues in the future. While many of these issues will be driven by changing substantive problems – increasing globalization, shifting economic and military power, emerging issues, and so forth – rational choice will be judged by how it addresses the theoretical and empirical questions that they raise. To explore these prospects, I consider three important questions that face international relations in theory and practice.

**Incorporating dynamics and change**

Rational choice might seem ineffective for studying change. The concept of equilibrium is inherently static since it is defined as the absence of any tendency to change. And the standard way to model “dynamic” choice is by redefining it as a static choice of an optimal strategy for all time, typically under an assumption of stable preferences (Kreps, 1990). Thus, even game models that represent sequences of choices through time, whether in the extensive form or as repeated play of a normal form game, typically take the fundamental structure of the situation as fixed and then focus on determining the equilibrium outcome.

Change is usually introduced through comparative static analysis of how the equilibrium shifts in response to exogenous change. The actual dynamic process and time path are not described, but bracketed under the assumption that actors adjust to a new equilibrium as it emerges. This failure to engage the process and mechanisms of
change becomes even more glaring when there are multiple equilibria so that choice among equilibria also needs to be explained. This is a particular shortcoming of the cooperation literature where the central question of when or how there will be a transition between equilibria – from anarchy to cooperation, or from one cooperative arrangement to another – is thereby ignored. Similarly, for historical analysis, it leaves open the question of when the prevailing equilibrium path will persist for reasons of path dependency or when there will be a shift to a superior time path.

There are some examples of more properly dynamic IR rational choice analysis. Power-transition theory (Gilpin, 1981; Organski and Kugler, 1980) argues that rapid shifts in power caused, for example, by differential economic growth will make war more likely. Powell (1999b) proposes a formal model that emphasizes the informational problems. By decomposing the concept of power shift, he shows that the size of a shift affects the probability of war but that its speed does not. James Fearon (1997) deepens this analysis by making shifts in power endogenous as concessions between states affect their future power balance. His finding that such changes do not lead to war (except in specifiable and unlikely circumstances) is an excellent example of a nonobvious result that could only be found through formalization.22

This sort of change, while important, is only partial. The reason is that the underlying substantive interpretation is deeply rooted in realism. Realism is fundamentally a theory of stasis that is premised on an enduring international anarchy with states as the primary actors.23 Rational choice offers a number of analyses of change within this overarching “anarchy” equilibrium. Change is typically driven by shifts in the distribution of power; adjustment mechanisms include war, changing coalition patterns (Niou et al., 1989; Wagner, 1986), or decisions to acquire arms (Intriligator and Brito, 1984). Similarly, realist analysis of political economy incorporates change driven by a changing distribution of power – as in hegemonic stability theory (Gilpin, 1981). Although all of these analyses focus on change at one level, their underlying presumption is that the overall system is stable. However, this restriction on the scope of change inheres in substantive assumptions drawn from realist theory rather than in the rational choice approach.24

Rational choice analyses of cooperation and institutions introduce the possibility of broader change in international politics.25 The cooperation literature establishes the possibility of attaining a different equilibrium than envisioned by realists. Institutions are seen as facilitating the attainment of cooperative equilibria and reinforcing and stabilizing those equilibria over time. Important mechanisms include improving information about potential joint gains from cooperation, reassuring actors that others also intend to join in the cooperative equilibrium, and providing timely information about behavior to diminish incentives to cheat. The impact cumulates within actors as they learn about the world and about each other, and changes their individual expectations regarding each other’s preferences and behavior. It also takes place at the collective level through the creation of norms of behavior and shared beliefs about the new equilibrium outcome. Here, rational choice moves beyond its individualism since equilibria and common knowledge are properties of the collective. Through this, institutional analysis is developing an analysis of institutions as independent factors in international politics and, possibly in some cases, as autonomous actors.

The foregoing is only a sketch of the past and evolving trajectory of rational choice institutionalism. My purpose is to show that rational choice potentially encompasses a wide range of change in international politics, although it has not fulfilled this broad agenda by any means. How successfully it can meet the constructivist challenge of explaining fundamental change in actor preferences (and identity), as well as change in the international political settings, remains
an active research question. My claim is only that it has the potential to do much in this regard.

Before turning to some (slightly) more detailed considerations of actor and preference change, let me distinguish two reasons why IR rational choice analysis has been reluctant to take on such issues. The first is because of substantive assumptions that are used in specific arguments but are not intrinsic to rational choice. Most realist analyses and some institutionalist analyses reject the assumption of fundamental change by maintaining a substantive emphasis on states as the central actors with constant goals and facing problems analytically similar to those of the past. Other analyses accept the notion of change both through institutions and within actors, but tend to see it operating more incrementally. This goes to a second methodological reason for reluctance to incorporate change, which essentially goes back to the revealed preference problem. Because considerations of change introduce even more flexibility into rational models, and preferences cannot be directly observed, the concern is that explanation will reduce to tautology. To avoid this, rational choice imposes limits on itself, such as fixed preferences. But in doing so it also limits its ability to examine possibly important phenomena. The next section considers some possible strategies for improving its analysis of changes in actors and in preferences.

**Endogenous actors and preferences**

Rational choice has traditionally assumed that the actors and interests are fixed in any analysis and has explained change in terms of changing constraints. The reason is that preferences are impossible to observe directly, whereas constraints are usually more observable. Under these conditions, fixed preferences allow for a tight analysis of many issues in an empirically falsifiable way, whereas assumptions of changing preferences lead to slippery and untestable arguments. However, the fixed preference assumption is not always valid and cannot handle all problems. In an article that is widely cited to justify the fixed preferences assumption, Stigler and Becker (1977) assert “not that we are clever enough to make illuminating applications of utility-maximizing theory to all important phenomena … [but that] no other approach of remotely comparable power and generality is available.” Their central argument is that fixed preferences *sometimes* provide a powerful analytic premise, whereas “assumptions of unstable tastes … really have only been ad hoc arguments that disguise analytical failures.”

Even this position – or at least the conventional interpretation of it – is overstated. Becker (1992, 1996) has moved well beyond it to consider “the evolution of preferences out of past experiences [which] seems far more intuitive, even when extended to institutions and culture, than the opposite assumption so dominant in economics that preferences are independent of the path.” These two positions – which Becker reconciles through the notion of stable meta-preferences – reflect the tension between sound methodological strategies and addressing certain, substantively important problems. IR rational choice analysis faces the same tension between maintaining the power of the approach while expanding its scope of coverage. Here I briefly sketch several different ways in which it can do so.

While treating states as aggregate actors with well-defined preferences is often a troublesome assumption for rational choice analysis, enriching the assumption provides a useful window for analyzing preference change. In IR, this has occurred as part of an effort to create a richer understanding of “state” motivations by unpacking them into their domestic components, both theoretically and empirically (Milner, 1998; Walsh, 2001). A wide range of examples includes analyses of political and economic coalitions (Frieden, 1991; Milner, 1988; Moravcsik, 1997; Rogowski, 1989), the analysis of two-level games (Evans et al., 1993; Putnam,
RATIONAL CHOICE AND INTERNATIONAL RELATIONS

1988), principal–agent models (Downs and Rocke, 1995), audience costs (Fearon, 1994), domestic commitment problems (Drezner, 2001), and macro-historical explanations of the evolution and changes in state preferences (Rosecrance, 1986; Spruyt, 1994). The logic is that aggregate preference change can be explained in terms of the changing relative influence of underlying constituencies, and their coalitional behavior, even where the preferences of the underlying constituencies are stable. Preference change can also result from variations in the regime type that determine how groups are aggregated (Goemans, 2000) or, within regime type, by variations in the specific aggregation rules (Rogowski, 1999). Such derivations of aggregate preferences impose strong theoretical and empirical requirements, but they offer a way to specify state preferences independently of observed state behavior.

A different type of preference change occurs when new actors with different goals enter an issue. Globalization, for example, may empower new actors ranging from activists to experts to firms. Insofar as they can be treated as goal-seeking actors, rational choice can explore the implications of their addition to the international setting. Of course, the heterogeneity of these new actors dramatically increases the complexity of the problem being analyzed and makes modeling more difficult. For that reason, analysis will lean toward soft rational choice informed by formal results. Examples of introducing non-state actors include bankers (Lipson, 1986), ship owners (Mitchell, 1994), environmental NGOs (Raustiala, 1997), soldiers (Morrow, 2001), airlines (Richards, 2001) and international business firms (Mattli, 2001). Indeed, even Keck and Sikkink’s (1998) constructivist analysis of transnational activists treats them as strategically rational agents at the international level. While this illustrates the complementarity of constructivist and rational approaches in examining the impact of new actors, there is no reason in principle why rational choice explanations may not be valuable in explaining the emergence of these nonstate international actors through the same logic used to explain the rise of domestic interest groups (Hansen, 1991; Schattschneider, 1935).

A third case is that of deliberate decisions to change actors and their preferences. This occurs when the members of an international arrangement, and therefore the preferences of the relevant actors, are objects of choice. Rational choice has had some preliminary empirical success in addressing membership within international institutions as a solution to an institutional design problem (Downs, et al., 2000; Koremenos, et al., 2001c). The new members can be states, but they also can be firms or NGOs admitted to participate on some international problem (Raustiala, 1997) or IOs that are either created or empowered on some issue (Abbott and Snidal, 1998; Thompson, 2001). In other cases, the membership is fixed but some actors seek to change the nature of other actors. Examples include human rights activists who seek internal reform in repressive states, international financial institutions that require program beneficiaries to adopt market reforms, or organizations such as the EU and NATO that impose “democracy” requirements on prospective members.

Finally, actors sometimes deliberately seek change in their own national “preferences.” This is what free trade coalitions hope will be the case as noncompetitive industries decline and what political leaders seek when they bind themselves to international arrangements to solve commitment problems (Drezner, 2001). States also engage in processes ranging from joint scientific research to cultural and political exchanges in order to learn more about the world, about each other, and about their joint expectations. By changing not only their own information but also their shared beliefs with respect to equilibrium behavior, they change their preferences over courses of action even if their underlying preferences are stable. Again, if the processes strengthen the relative position of some domestic actors, that may also change aggregate state preferences. More dramatically,
states make rational plans to join international institutions such as the EU, WTO, or international legal agreements with the understanding – often reflected in resistance from some quarters – that those institutions will significantly shape their own future course. Rational choice has a largely unfulfilled role to play in analyzing this sort of deliberate preference change.

The examples illustrate different strategies for introducing preference change through aggregation processes, the introduction of additional actors, through the choice of actors, or through changing information and the external institutional environment. They are speculative, although in each case some progress has been made. But these approaches also raise tensions between the advantages of parsimony and the value of empirical richness, and between formal and softer analyses. Finally, despite these possibilities, rational choice (and for that matter, other approaches) should remain modest about the extent to which it can offer systematic analyses of the complicated problem of actor and preference change.

Normative and policy analysis

Because the research impetus of rational choice has been heavily “scientific” such that it is even labeled “positive” theory, both proponents and critics often see it as far removed from normative analysis. This is ironic since rational choice began as a normative enterprise (Stein, 1999) and lends itself readily to normative analysis, at least along the utilitarian lines from which it developed. So it is fair to say that rational choice in IR has not fulfilled its normative potential, but not fair to say that it cannot do so (Wendt, 2001).

Rational choice already contains important normative elements. At an individual actor level, it is implicitly a theory of how people should behave – what is the best choice? – in a given situation. At a collective level, it examines how groups can do better through cooperation to attain a superior equilibrium. But while these efficiency considerations have important normative content – and, to reiterate, efficiency need not be defined in terms of material interests – rational choice can and should engage other normative considerations.

Distributional questions are slowly emerging as an important area of inquiry in IR rational choice and bring both a positive and normative dimension. Theoretical work has moved beyond the use of coordination problems to exemplify “the” distributional problem to a realization (especially via the folk theorem) that distributional issues are pervasive. An important literature has emerged exploring how distributional differences inhibit cooperative efficiency gains (Fearon, 1998; Krasner, 1991; Morrow, 1994a). Empirical work has begun to examine the impact of cooperative arrangements on distributive outcomes at both the international and domestic levels (Goldstein, 1996; Gruber, 2000; Oatley and Nabors, 1998). By emphasizing that there may be losers as well as winners from “cooperative” schemes, analysis of this sort invites a normative evaluation of the achieved equilibrium in comparison to alternative possibilities.

Institutional design questions bring rational choice yet closer to normative analysis. Analyses of institutional arrangements have addressed how issue linkage (Mitchell and Keilbach, 2001; Sebenius, 1983), membership (Koremenos et al., 2001c; Pahre, 2001), and incorporation of escape clauses and other forms of flexibility (Koremenos, 2001) can remedy distributional impediments to cooperation. Since institutional analysis is premised on the analysis of alternative equilibria, it is a short step from asking what institutions states will design to asking what institutions they should design. For example, bargaining models typically predict the outcome among individuals in particular circumstances but can equally be used to ask what protocols and institutions will lead to preferred outcomes. Brams and Taylor (1996) illustrate the possibilities in their investigation of how to devise fair allocation devices among rational agents. IR has only begun to engage this type of analysis, but rational choice is
eminently suited for it and already offers a large number of insights (Young, 1994).

It is less clear how fully rational choice can incorporate other normative issues – such as justice, appropriateness, or autonomy – but it can certainly contribute to an analysis of their impact in IR. Of course, insofar as other values can be treated as objectives of actors, rational choice can readily incorporate them. Despite the danger of trivializing other values by simply relabeling them as interests, this can be valuable for understanding their impact (and perhaps the impact of the actors who carry them), as well as the possibilities of and strategies for attaining them. Finally, by showing that good institutional design must be consistent with interests to be effective, rational choice can help normative analysis avoid foundering on naïve idealism.

Rational choice can further benefit by addressing the “positive” impact of normative conceptions by way of their influence on actors’ behavior. This is one of the central insights raised by constructivist critiques of interest-based approaches. Rational choice often has nodded implicitly to it through reference to the role of normative considerations in creating “focal points” to guide choice among multiple equilibria. But no adequate theory of focal points – normative or otherwise – has been developed.

Another area where normative considerations appear important is in international legalization. Rational choice has now begun to explain the form and content of legalization (Goldstein et al., 2000). While significant insights have been generated in seeking to explain legalization from a strictly rationalist perspective (Goldsmith and Posner, 1999, 2002a, 2002b), a purist approach may overlook the normative force of law reflected in such key concepts as legitimacy, obligation, or compliance pull (Abbott and Snidal, 2000, 2002; Hurrell, 1993; Finnemore and Toope, 2002). Because these concepts are notoriously vague, however, rational choice can play an important role in tightening the theoretical discussion and sorting out the logic of the claims. An important example is the debate on compliance initiated by the Downs et al. (1996) critique of “managerial approaches” (Chayes and Chayes, 1995) and its extension into a more general critique of the legal “transformationalist” school (Downs et al., 2000). However, the goal should not be to show how rational choice can make the normative perspective redundant, but to explain how rational incentives and normative understandings interact through legalization.

A turn to normative questions also suggests a need to shift the emphasis of tools within the rational choice toolkit. Primary concern for problems of enforcement and associated problems of commitment and information led to an emphasis on noncooperative game theory as the appropriate method of analysis. While enforcement remains an important issue, it is not the only one. Other models that de-emphasize enforcement problems may be more valuable for these new problems. Cooperative game theory with its focus on how the gains from cooperation are divided may provide better insights into distributional issues. Similarly, attention to questions of mechanism design – that is, how to design institutional arrangements so as to achieve the desired ends – may be more relevant as the field shifts to address normative issues. Now, instead of using rational choice to ask the positive question “Given actors’ interests and (institutional and other) constraints, what will be the outcome?,” we turn the analysis around to ask “Given actors’ interests, how should we rearrange institutional constraints to achieve our normatively desired outcome?” Rational choice analysis in IR is barely on the threshold of asking such questions, but they are the means through which its normative and policy potential can be unleashed.

**CONCLUSION**

While the analysis of IR in terms of goal-seeking agents has long been a central part of IR theory, rational choice analysis has
become more explicitly theoretical by drawing upon both formal and informal analytic results. Its considerable successes have led to optimism and even arrogance regarding its future prospects, but there have also been failures and shortcomings which critics appropriately have been quick to point out. Rational choice cannot resolve all of these challenges, but by addressing them seriously it can increase its own power to enhance our understanding of international affairs.

The view of rational choice adopted here is expansive. In particular, I did not draw a line between formal and soft rational choice because their relation is highly symbiotic. Formal results often provide the hard kernel behind softer analysis, while softer analysis encourages a wider range of interpretation and application of the model. For this reason, soft rational choice approaches lie behind many of the empirical tests of the theory because they facilitate the adaptation of the abstract theory to different issues, cases, and historical periods.

While rational choice has numerous shortcomings, some of its seeming failings are best understood either as important features of the world or as a result of appropriately cautious strategies for analyzing it. The existence of multiple equilibria, for example, may indicate not a failure of a model but an important indeterminacy in the world, and in our ability to predict outcomes. Of course, it remains a reasonable goal for rational choice – in concert or in competition with other approaches – to explain how and why a specific equilibrium is attained. Similarly, rational choice models that leave out important considerations such as change in preferences thereby restrict their ability to handle some important international issues. But this limitation may have a benefit in allowing other questions to be studied more carefully. Of course, rational choice should seek to expand its range of coverage with regard to excluded issues but, in doing so, it should remember that the power of its analysis rests partly on the limits it imposes. Without such limits, rational choice (like other approaches) explains nothing by pretending to explain everything.

Thus rational choice in IR should be neither defensive nor arrogant. It does not have to be defensive because it has led to significant advances in our understanding of international politics. Many of its limitations are self-imposed for good reasons, and others provide challenges that it can be expected to address. But rational choice should not be arrogant, because its critics have identified significant shortcomings in its theoretical and empirical work. Rational choice cannot answer all these challenges, and the answers it can give are often “soft” compared to its aspirations.

NOTES

I thank an anonymous referee, Susan Pratt, Alexander Thompson, and the editors of this volume for helpful comments on an earlier draft of this chapter.

1 For a representative range of critiques, not all specifically addressed to international relations, see Elster, 1989; Friedman, 1996; Green and Shapiro, 1994; Katzenstein, 1996; Monroe, 2001; Ruggie, 1998; Walt, 1999; Wendt, 1999; Yee, 1997.

2 For other recent efforts to look at this question from different perspectives, see Lake and Powell, 1999a, a number of the articles in Katzenstein et al., 1998, andHasenclever et al., 1997.

3 Fearon and Wendt (2002) discuss the relation between constructivism and rationalism.

4 As always, the word “theory” is a bit elusive here. Rational choice is essentially a normative theory of how actors should choose in different (abstract) circumstances that becomes a positive theory when actors are assumed rational and the empirical circumstances (e.g., of IR) are specified. Formalization introduces a parallel mathematical “theory” which provides an abstract representation with minimal substantive content. A particular empirical interpretation of this abstract theory turns it into a substantive theory.

5 I thank Thomas Risse for pressing me on this point.

6 Here and elsewhere I use specific references as iconic representations of work in the field. There are too many relevant contributions to provide an exhaustive bibliography.

7 Axelrod’s work on the evolution of cooperation (1984) and agent-based models (1997; see also
Cederman, 1997) are illustrative. See also Rubinstein’s (1998) effort to model bounded rationality and Fearon and Wendt (2002).

8 Even differences within rational choice methodology – for example, over the use of one solution concept rather than another – are ultimately debates over how the actors would (or should) handle a particular problem.

9 Powell (1999b) and Lake and Powell (1999b) provide a good overview of important issues pertaining to the use of models in international relations. See Walt (1999), for a critique and the subsequent responses by Bueno de Mesquita and Morrow (1999), Martin (1999), Niou and Ordeshook (1999), Powell (1999a), and Zagare (1999).

10 Allan and Dupont (1999) provide an excellent overview of the diversity of alternative models. They develop a typology of 16 different types of social model, of which they categorize 8 as rational. They also pay special attention to the relation between model choice and empirical applicability.

11 The “folk theorem” says that any feasible payoff combination can be an equilibrium in an ongoing interaction provided that the players are sufficiently patient (i.e., don't discount the future too heavily) and that every player receives at least as much as they can guarantee themselves even if everyone else gangs up on them. An important consequence is that most such interactions will have “multiple equilibria.”

12 See Hasenclever et al., 1997, for an excellent overview of this debate.

13 A clear success is the introduction of “sub-game perfection” to eliminate incredible threats, while concepts such as sequential or Perfect Bayesian equilibria have also been widely used to good effect. But the value and empirical applicability of more complicated equilibrium “refinements” has been increasingly challenged (Allan and Dupont, 1999).

14 This does not mean that the actor must be able to “solve” the model or understand it in its entirety, but it must be able to reasonably perform the tasks assigned to it. For example, an international negotiator need not understand the Arrow (1951) theorem in order to recognize opportunities for strategic agenda setting or linkage.

15 Well-known deviations include the bureaucratic and psychological schools of decision-making, especially the literature which builds on work such as that by Kahneman and Tversky (1979) in challenging rational choice directly (Levy, 1997). Some of the internal rational choice critiques of key concepts such as discounting (Laibson, 1997) and risk (Rabin and Thaler, 2001) also suggest limits inherent in the logical theory as a representation of rational decision making.

16 Without diverting my argument that rational choice can meet the challenge of testing head-on, the value pf curve-fitting in IR should not be dismissed. Establishing systematic empirical relations is a real contribution to a field that remains fairly short of clear and systematic facts. The ‘democracies don’t go to war’ correlation, for example, has stimulated important work on both the theoretical and empirical fronts. Even identifying a situation as fitting a particular model (e.g., as prisoners’ dilemma) may enhance our understanding of it.

17 See not only the standard cite to Stigler and Becker (1977) but Marshall’s (1910) early discussion of preferences and time periods.

18 For a critique of the empirical contributions of rational choice, including the expected utility theory of war, see Walt (1999), and replies by Bueno de Mesquita and Morrow (1999), Martin (1999), Niou and Ordeshook (1999), Powell (1999), and Zagare (1999).

19 The most controversial inclusion here is (deliberately) the work of Walt, who has been highly critical of rational choice and especially formal models. See Frieden, 1999: 50–1 for a cogent analysis of Walt’s work as fitting in the rational choice tradition – or at least of how it could be improved by a more self-conscious attention to the (soft, not formal) rational choice approach.

20 Consider also examples such as Goldstein (1993), and Goldstein and Keohane (1993), which are critical of rational choice as too limited (by leaving out ideas) but proceed in a largely complementary manner; works such as Spruyt (1994) that engage important elements of rational choice arguments but move significantly beyond them; or works such as Pape (1996) that use a rational choice framework to sort out different verbal arguments.

21 Analytic narratives can be seen partly as an elaboration of earlier discussions of the use of rational choice approaches in “interpretive” accounts (Ferejohn, 1991; Johnson, 1991). Indeed, the possibility of its strictly interpretive usage reminds us that rational choice has no necessary connection with positivism and “testing.”

22 Another example of more dynamic analysis that has been applied to IR is Axelrod’s (1984, 1997) analysis of the evolution of cooperation, which can be interpreted from a rational choice as well as other perspectives. It shows how dynamic processes of selection, imitation, or learning may lead to changed outcomes.

23 Thus, Robert Gilpin’s (1981: 211) study of change maintains “that the nature of international relations has not changed fundamentally over the millennia.” See also Waltz, 1979: 65–6, and Krasner, 1999, with respect to sovereignty.

24 Of course, the realist substantive position may be right. See Grieco, 1997, for a persuasive case for the value of studying the constant factors in international life across time and systems.

25 For a full treatment of international institutions, see Simmons and Martin’s contribution to this
volume. Here, I offer more of a speculative sketch of the trajectory of rational choice approaches.

26 Choice over time also raises deep problems of inconsistency even when preferences are constant (Strotz, 1955–56), which are further aggravated by changing preferences (Hammond, 1976). One could interpret such results as indicating the inability of rational choice to deal with changing preferences over time. In my view, a superior interpretation is that the models reveal deep problems relevant to any analysis of changing preferences. See Schelling (1978) and Elster (1979), for extensions of these problems that bridge beyond rational choice, and Becker (1996), for an effort to incorporate them within rational choice.


28 Empirical work, often but not always closely guided by rational analysis, has taken the lead, but Milner (1998), discusses how theoretical rational choice can catch up with it.

29 The standard distinction between cooperative and noncooperative games is that agreements are binding in the former and not in the latter. Moulin (1995) argues that this misunderstands the self-enforcing nature of the core and other stability concepts in cooperative game theory. More important, for some IR problems – such as agreements within the EU or many agreements among the OECD countries – there is no more doubt that parties will abide by their agreements than in domestic contexts. In such cases, cooperative game models can allow for a tighter focus on the bargaining–distribution problem by de-emphasizing the enforcement problem.

30 I thank an anonymous reviewer for suggesting that I mention this literature. For an overview of mechanism design, see Myerson (1991) and Mascielli et al., 1995; and, for a less technical introduction, see Dutta (1999).

REFERENCES


