

Trustees of Princeton University

---

The Game Theory of International Politics

Author(s): Duncan Snidal

Source: *World Politics*, Vol. 38, No. 1 (Oct., 1985), pp. 25-57

Published by: Cambridge University Press

Stable URL: <http://www.jstor.org/stable/2010350>

Accessed: 22-06-2016 04:35 UTC

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at  
<http://about.jstor.org/terms>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



*Cambridge University Press, Trustees of Princeton University* are collaborating with JSTOR to digitize, preserve and extend access to *World Politics*

# THE GAME *THEORY* OF INTERNATIONAL POLITICS

By DUNCAN SNIDAL\*

THE application of game theory to international politics is hardly new, but there has been a recent increase in the popularity of the approach. This resurgence has been associated with new applications of game models to international political economy in addition to their now standard role in military-political strategic analysis. This is a timely antidote to the exclusive usage of strategic analysis to refer to military affairs. What is fundamental to strategic analysis is not the specific subject matter of military or economic issues, but a basic conception of how we understand politics among states. This conception of nation-states as interdependent, goal-seeking actors lies at the heart of strategic game analysis; it is applicable across different issue areas.

The ultimate promise of game theory lies in expanding the realm of rational-actor models beyond the restrictive confines of the traditional Realist perspective to a more complex world where concern is less exclusively with problems of conflict and as much with problems of cooperation. This expansion will provide a stimulus to the ongoing theoretical project of integrating the military and political-economy sides of international politics. One important aspect of this theoretical integration is the development of models capable of encompassing different issue areas that are usually treated as fundamentally disparate. Individually, the articles in this volume advance that goal by applying game models across a wide range of international political, military, and economic issues. Collectively, the articles show the value of game theory in a broader theoretical enterprise: understanding different questions of international politics within the same theoretical framework. Thus, as will be elaborated upon throughout this article, the ultimate payoff of game theory is the use of game *models* to understand different aspects of international politics in terms of a unified *theory*.

A second valuable aspect of this resurgence of interest in game theory

\* Raymond Duvall has provided a number of important ideas on how to formulate the argument of the paper. I have also benefited greatly from discussions with the other participants in this symposium, and from comments by Kenneth Oye, Robert Axelrod, and especially Robert Keohane. Chris Achen and John Freeman made valuable comments on earlier drafts, as did participants in the Program on Interdependent Political Economy (PIPE) Workshop at the University of Chicago.

is a greater concern for empirical application of the models. Earlier military-strategic analyses were based on deductive arguments, but remained untested beyond anecdotal embellishment. Perhaps this could be excused by the fortunate fact that there had been insufficient data on the outbreak of nuclear war as the central dependent variable. However, relevant historical evidence on non-nuclear deterrence also has only recently been systematically exploited.<sup>1</sup> One virtue of recent game theory applications to issues of both military and political economy is its attention to the empirical implications and evaluation of the deductive theory. In examining the usefulness of game theory for theorizing about international politics, I will emphasize the requisites for this empirically oriented side of the game-theoretic enterprise.

Applying game theory to a substantive body of knowledge such as international relations raises a host of difficult empirical questions. For example: Who are the relevant actors? What are the rules of the game? What are the choices available to each actor? What are the payoffs in the game? Is the issue best characterized as single-play or repeated-play? In analyzing any particular issue, it is impossible to answer all of these questions with certainty. Game theory often seems to demand more information than can feasibly be obtained. Ironically, it cannot always adequately incorporate other important available information—including relevant historical details about the context of interaction, insights into the personalities and behavior of decision makers, and understandings of the diplomatic or foreign policy process. These shortcomings of game theoretic analysis have led some analysts to conclude that its usefulness as a theoretical guide to the empirical study of international politics is seriously impaired.

This conclusion shows a misunderstanding of the power of (game) theory by treating it as a descriptive and not as an analytical tool. Too many “applications” of game theory have merely been in the spirit of sorting out whether the Cuban missile crisis was really Chicken or Prisoners’ Dilemma. Such usage may be helpful for reconstructing and interpreting particular events, but it misinterprets the primary value of game theory as that of redescribing the world, and is therefore limited as a test of game theory. It would be a more appropriate test of a

<sup>1</sup> Among more recent works, see Alan Alexandroff and Richard Rosecrance, “Deterrence in 1939,” *World Politics* 29 (April 1977), 404-24; Bruce Bueno de Mesquita, *The War Trap* (New Haven: Yale University Press, 1981); Alexander George and Richard Smoke, *Deterrence in American Foreign Policy* (New York: Columbia University Press, 1974); Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976); John Mearsheimer, *Conventional Deterrence* (Ithaca, NY: Cornell University Press, 1983); and Glenn Snyder and Paul Diesing, *Conflict Among Nations* (Princeton: Princeton University Press, 1977).

deductive theory to investigate the empirical correctness of its analytical predictions. This requires giving empirical content to it through its assumptions (e.g., about preferences and payoffs) rather than just adapting the theory (or one of its many models) to fit some historical or current event.

The distinction can be illustrated by an example of game theory used as a descriptive rather than an analytical tool. In *Conflict Among Nations*, which is widely regarded as an ambitious application of game theory to international relations, Snyder and Diesing use game models of 16 historical cases to investigate bargaining in international crises. Their technique is to reconstruct the game structure underlying each crisis from a detailed historical analysis of events. They show a sophisticated awareness of the difficulty of such reconstruction and would not be surprised if (as often happens in this sort of analysis) others offered different interpretations of various crises. Such a challenge, however, is only a test of their skill in reconstructing a crisis in game terminology, and not a test of the theory itself. Indeed, because Snyder and Diesing use the totality of the crisis (including the outcome) to generate the descriptive game model, their use of game theory does not produce any predictions that could be empirically falsified. None of the deductive power of game theory is employed. Thus, their work is not an example of the empirical application of game theory even though it illustrates the purely descriptive use of game models.<sup>2</sup>

The real power of game theory, for both empirical and theoretical purposes, emerges when it is used to generate new findings and understandings rather than to reconstruct individual situations. This objective requires treating game theory as a *theory* of international politics in terms of the goal-seeking behavior of states in an interdependent international system. It means that game models will not be useful in predicting outcomes that are largely determined by nonpurposive or nonsystematic behavior. If all or most cases are determined by such factors, then game theory will have little to say about international politics. If the underlying assumption of self-interested action by strategically rational states is correct, however, and preferences, strategies, and payoffs can be modeled accurately, then the theory will generate important testable predictions.

<sup>2</sup> My comments are not intended as criticism of an ambitious and insightful enterprise. The authors' descriptive use of game theory is appropriate, given an ultimate concern with different models of decision making and bargaining rather than with game theory, which is used only in "the limited role . . . [of] depicting the structure of a crisis" Snyder and Diesing (fn. 1), 87. Nevertheless, if any analytical game theory approach were attributed to the work, a game theorist could reasonably object that it had been emasculated (cf. p. 182).

This position is based on the now widely accepted view that the purpose of any theory—including game theory—is not to reproduce reality, but to increase our understanding of fundamental processes by simplifying it. For this reason, it is not desirable to incorporate all of the details of any individual case. Simplicity and abstraction guide us through a morass of information to focus on more fundamental issues.

While the simplicity of game models leads to a clarity that illuminates social phenomena, the deductive apparatus of game theory allows us to infer new understandings about international politics. The best-known example is Prisoners' Dilemma. Analysis in terms of this  $2 \times 2$  game provides insights on issues such as GATT or SALT, which could never be achieved by, say, archival research alone. Expanding the analysis to N-person Prisoners' Dilemma and the logic of collective action explains why international cooperation sometimes fails even when it is in the interest of all states. Further investigation of the impact of states of different sizes and of hegemonic "leadership" (in terms of privileged groups) shows how various distributions of capability and interest facilitate or impede collectively optimal outcomes.<sup>3</sup> And changed incentives in dynamic games explain why cooperation has been more successful through GATT than through SALT, where mutual suspicions reinforce perceptions of substantial danger if the other side were to renege on cooperation.

These are the fertile directions in which game theory can lead empirical investigation of international politics. In the first section of this paper, I elaborate on game theory as a theoretical approach to international politics by contrasting it with metaphorical and other uses of games. This exercise lays a foundation for discussing the requisites for empirical application of the theory. In the second section, I discuss the overarching theoretical constructs in game models and their implications for applying the theory to specific international issues. These constructs raise a new set of questions that expand our understanding of particular issues. In the third section, I examine theoretical extensions of simple game models that expand the contextual richness of empirical applications by incorporating key features of international political interactions. The discussion in both of these sections makes reference to the technical game literature, but the focus is on the empirical issues of international politics raised by game-theory approaches.<sup>4</sup> In the final section, I sum-

<sup>3</sup> Duncan Snidal, "The Limits of Hegemonic Stability Theory," *International Organization* 39 (Autumn 1985), 579-614.

<sup>4</sup> The articles in this volume presuppose a basic familiarity with game-theoretic approaches. A good introduction is Henry Hamburger, *Games as Models of Social Phenomena* (San Francisco: W.H. Freeman, 1979); Martin Shubik, *Game Theory in the Social Sciences: Concepts*

marize the usefulness and limits of game theory as a theory of international politics and as a guide to empirical research.

## I. GAME THEORY AS A THEORY OF INTERNATIONAL POLITICS

The diverse logics of applying game theory to empirical phenomena are highlighted by its different applications as metaphor, analogy, model, or theory. Even though these terms are sometimes used interchangeably, distinguishing among them illuminates the relation of the game approach to empirical evidence. These differences go well beyond hackneyed calls for more rigorous approaches even if some uses are inherently more rigorous than others. Instead, different logics in using game theory reflect different understandings of how to apply it empirically. Most applications of game theory to international politics have been in terms of metaphor and analogy; they are unsatisfactory guides for application in terms of model and theory. Since we are concerned here with empirical application of deductive rather than purely descriptive models, these distinctions are important. Finally, in presenting these categories, the object is not to provide definitive definitions of terms, but to demonstrate the varying implications that the different logics hold for game-theoretic approaches to international politics.

### A. METAPHOR

The use of metaphor is pervasive not only in literature, but in science and policy. Mechanical and biological metaphors are common in the social sciences. An applied example is the metaphor of “falling dominoes,” which has been significant in postwar U.S. foreign policy. The power of the metaphor rests on its loose and open-ended nature. It invites us to speculate and engages us in creative conjecture. No formal deductive apparatus is involved: an implied comparison of two entities is used to infer further properties or conclusions from one to the other.<sup>5</sup>

Metaphors allow for the creative transfer of ideas across intellectual realms. This explains their heuristic and expository value—especially in exploratory stages of theory construction where their open-ended character is amenable to innovative thinking. However, they are also highly susceptible to misunderstanding and misuse. Open-endedness can de-

---

*and Solutions* (Cambridge: MIT Press, 1983) provides a more advanced treatment. Two accessible “classics” are R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: Wiley, 1957), and Thomas Schelling, *Strategy of Conflict* (Cambridge: Harvard University Press, 1960).

<sup>5</sup> Martin Landau, “On the Use of Metaphor in Political Analysis,” *Social Research* 28 (Autumn 1961), 331–53.

volve into sloppiness or silliness if metaphorical transfer is (mis)applied without thorough attention to the correspondences in the properties and conclusions of the entities being compared. These dangers are well illustrated by the more “vulgar” and misleading applications of biological and mechanical metaphors to understanding society.

Thus the use of metaphors in social science needs to be somewhat guarded; metaphorical richness must be progressively restricted by more precise formulations as research advances. For example, consider the Hobbesian state of nature as a fundamental metaphor for international politics. Although instructive for highlighting the security dilemma among states, it is misleading when it is overinterpreted to the conclusion that any international cooperation requires an international government comparable to the Hobbesian sovereign. This confusion does not arise from a more sophisticated specification of the metaphor, which recognizes salient differences between the two political environments (e.g., states are inherently more secure in the international system than are individuals in the Hobbesian state of nature) that allow for decentralized international cooperation.<sup>6</sup>

Increased rigor of specification is not to be confused with expression in a different form—specifically with mathematical as opposed to purely verbal statement. We do not improve a metaphor simply by translating it into a game matrix. Glib assertions that “Issue *X* is Prisoners’ Dilemma,” or that the “Cuban missile crisis was a game of Chicken,” efficiently convey a metaphor, but do not make the metaphor more plausible or take much advantage of the power of game theory. Typically, such statements simply restate what we already know—perhaps embellished by means of a particular reconstruction of historical events. Real rigor requires tightening the correspondence between the metaphor and the issue at hand. Analogies and models are to some extent simply more controlled metaphors, although they are further distinguished by their respective logics of inference.<sup>7</sup>

<sup>6</sup> The Hobbesian (domestic) state of nature is contrasted with the current international situation in Charles Beitz, *Political Theory and International Relations* (Princeton: Princeton University Press, 1979). Decentralized international cooperation is dealt with from a strategic-actor perspective in Robert O. Keohane, *After Hegemony* (Princeton: Princeton University Press, 1984), and from a more purely game-theoretic perspective in Duncan Snidal, “Interdependence, Regimes and International Cooperation” (Ph.D. diss., Yale University, 1981). See also the related articles in Stephen Krasner, ed., *International Regimes* (Ithaca, NY: Cornell University Press, 1983).

<sup>7</sup> Metaphors are distinguished by the looseness of their correspondence rules, but *not* necessarily by the absence of mathematical sophistication. For an example of a mathematically sophisticated (though ultimately nonfruitful) metaphor, recall Paul Smoker’s use of the harmonic motion of springs as a metaphor for arms rivalry, in “The Arms Race as an Open and Closed System,” *Peace Research Society (International) Papers* 7 (1967), 41–62. For a

## B. ANALOGY

The key to analogy is the special type of reasoning it invokes. On the basis of certain closely established resemblances between two entities, further similarities are inferred. If  $X_i$  represents the  $i^{\text{th}}$  property of entity  $X$ , an analogy between  $A$  and  $B$  is established by demonstrating a correspondence between a number (say,  $j$ ) of their properties and/or relations ( $A_1 \leftrightarrow B_1, A_2 \leftrightarrow B_2, \dots, A_j \leftrightarrow B_j$ ). On the basis of the analogy, we infer (tentatively) that for properties  $A_{j+1}, A_{j+2}, \dots$  (etc.) known to pertain to  $A$ , there exist corresponding properties  $B_{j+1}, B_{j+2}, \dots$  (etc.) in  $B$ . This inferential reasoning is external to both entities and resides in the comparison between them. Neither side of the analogy need contain any internal deductive structure (i.e., none of the properties  $j+1, j+2, \dots$  (etc.) can necessarily be logically derived from properties  $1, 2, \dots, j$  for either  $A$  or  $B$ ). The logic of inference is then heavily inductive, resting on observed isomorphisms and, ultimately, on empirical confirmation of derived propositions (i.e., properties  $j+1, j+2, \dots$  etc.).

Analogic inferences are tentative until empirically confirmed. A central task of analogic reasoning is to distinguish *negative* (i.e., incorrect) from *positive* (i.e., correct) analogies. *Neutral* analogies, not yet known to be correct or incorrect, are viewed agnostically until more evidence is available. Therefore, analogy is a guide to empirical observation that points out parallels between properties of different phenomena, rather than a deductive tool whose conclusions about the existence of those properties we can subscribe to without question. Its main difference from the metaphor is the tighter specification of correspondences between properties and the closer evaluation of conclusions to which this leads.

One fashionable way of using analogy has been to relate international politics to neoclassical microeconomic models. Central among these has been the analogy between states in the international political system and firms in an oligopolistic market (and, implicitly, the game-theoretic structure underlying that market). This analogy is constructed from a number of postulated empirical correspondences, including:<sup>8</sup>

---

critical discussion, see Peter Busch, "Mathematical Models of Arms Races," an appendix to Bruce Russett, *What Price Vigilance?* (New Haven: Yale University Press, 1970).

<sup>8</sup> The oligopoly analogy is widespread. The correspondences presented here can be found in Kenneth Waltz, *Theory of International Politics* (Reading, MA: Addison-Wesley, 1979), esp. chap. 7. Other examples of microeconomic analogies are Robert Gilpin, *War and Change in World Politics* (New York: Cambridge University Press, 1981) and Robert O. Keohane, "The Demand for International Regimes," *International Organization* 36 (Spring 1982), 325-55.



These correspondences establish the analogy that allows other inferences from oligopoly theory to the international system. They include the proposition that, since oligopolistic market concentration leads to market stability and fewer price wars, concentration of power in the international system will lead to system stability and less international conflict. The status of this inference is that of a hypothesis whose *a priori* plausibility depends on whether the initial correspondences are compelling. If empirically verified, the positive analogy becomes a new supporting correspondence which, in turn, buttresses the strength of the analogy as further implications are pursued. There may also be negative analogies, however, which must be identified as incorrect inferences. One example is the disanalogy between the role of elimination through competition which plays an important role in the marketplace but is inconsequential in the system of states. Another difference is that economic outcomes are evaluated primarily in terms of systemic properties such as efficiency rather than of the fate of individual firms, while international systems are evaluated fundamentally in terms of the impact they have on individual states.

Although analogy can be a powerful tool for conceptualizing international politics, its limitations are apparent. Disanalogies can be recognized but not eliminated; features of interest that have no corresponding property in the analogy cannot be analyzed. Finally, the logic of inference is primarily inductive, and no direct use is made of deductive logic. Since that is where the power of the game-theoretic approach lies, analogy uses only a portion of this power. A better alternative is to use game-theory models that directly incorporate the most salient features of the international system.<sup>9</sup>

There is an even stronger concern for a tight correspondence between an entity and its model. Because it can be abstract and “constructed,”

This content downloaded from 175.45.185.0 on Wed, 22 Jun 2016 04:35:32 UTC  
All use subject to <http://about.jstor.org/terms>

a model can be tailored precisely to the problem under investigation. Its key distinguishing characteristic is a formal logic that is both deductive and internal (as opposed to the inductive and external logic of analogy). Using the same notation, let  $A$  be a model of  $B$  with correspondences  $1, 2, \dots, j$  established. Property  $A_{j+k}$  is a conclusion of the model if it can be deduced from some combination of  $A_1, A_2, \dots, A_j$ . The corresponding property  $B_{j+k}$  is then asserted on the basis that it follows logically from the corresponding combination of  $B_1, B_2, \dots, B_j$ . This differs from analogy in that no recourse is made to induction beyond the initial establishment of the isomorphism. Instead, the model simply illuminates a logical relationship among the properties of  $B$ , which is more expeditiously demonstrated among the corresponding relations in its model  $A$ .

As a result, the model has a more closed form than the analogy. In an analogy, establishing correspondences  $A_i \leftrightarrow B_i, \dots, A_j \leftrightarrow B_j$  does not guarantee that each  $A_{j+k}$  implies a corresponding  $B_{j+k}$ . Some analogic inferences may be incorrect (and thus set aside) without significantly harming the analogy—provided that sufficient correspondences are established and that other correct inferences pertain. By contrast, the internal deductive argument of a model that allows us to conclude  $A_{j+k}$  based on properties  $A_1, A_2, \dots, A_j$  of  $A$  indicates that, *if our correspondences are valid*, then conclusion  $B_{j+k}$  *must* also apply. If empirical evaluation shows  $B_{j+k}$  does not hold, then the model itself must be reformulated. The result is a more closed form: negative analogies are not permitted to coexist alongside positive analogies. We are forced to revise and reconstruct the model.<sup>10</sup>

The compensating features of (empirically verified) models are that neutral analogies (i.e., untested conclusions) can be accepted until disproven, and that the logical structure of the model suggests new avenues of inquiry. Therefore correspondences must be tighter in a model than in an analogy. A wrong correspondence in an analogy need not impair the (external) logic of inference if sufficient other correspondences still

---

analogy, see Brian Barry, *Sociologists, Economists and Democracy* (Chicago: University of Chicago Press, 1978).

<sup>10</sup> I do not want to convey too pristine a view of how models are developed. Models are always constructed with an eye toward some of their inferences (what assumptions are needed to produce a certain conclusion). Moreover, the *ceteris paribus* clause is often invoked to deal with important correspondences that are *not* contained in the model. However, the logical structure of the model forces our theoretical assumptions and conclusions to be consistent and leads to other wholly unanticipated inferences. The *ceteris paribus* clause should not become a refuge from incorporating further considerations into the model; it should be a stimulus for its progressive refinement.

hold. In a model, a single wrong correspondence may disrupt the (internal) logic of deduction and produce false conclusions. Thus, with stronger requirements for its correspondence rules and stronger deductive powers, a model makes stronger claims about the world.<sup>11</sup>

Understanding and explanation is deepened by the deductive form of the model. The need for tightness in correspondences, and for revision of the model when it is inadequate, requires that our knowledge be more precise. Whereas in analogy it is unclear which correspondences (or how many) are necessary, the model clarifies which ones play a role in the logic of inference (i.e., those that enter into the deductive reasoning). The model thus promotes parsimony and simplicity. It also directs us to evaluate the accuracy of the most important correspondences, and allows us to investigate how changes in correspondences (assumptions) might lead to different conclusions.

Finally, models of processes or of things (i.e., of a specific empirical phenomenon such as a particular arms race) need to be distinguished from models of theories (i.e., of a general category of phenomena such as arms races). The former involve abstraction of an entity's properties to represent them in a simpler set of relations. This sort of model may incorporate a theory, but may also be consistent with many theories. A model of a theory, on the other hand, contains a set of linked law-like statements applicable to a range of phenomena. Game theory poses a healthy tension between these types of models: a game model may be treated as a straightforward representation of a specific phenomenon or as involving a broader theoretical perspective and a more general interpretation of the particular event.<sup>12</sup>

#### D. THEORY

Since the same representation can be a model for different theories, interpretation of a model depends on the theory in which it is embedded.<sup>13</sup> The theory contains a deductive structure plus an inter-

<sup>11</sup> Models are *not* primarily distinguished from analogies and metaphors by mathematical sophistication. (See note 7.) For example, physical models and analogue machines (including computer simulations viewed as physical machine representations) are models that are not in explicit mathematical form. Mathematics is simply a particular way of expressing a model. It is useful in forcing us to tighten up correspondences, in exposing weaknesses in a model or metaphor, and in providing a powerful means of pursuing deductive implications.

<sup>12</sup> Again, the distinction between a theory and its model is not always clear. Mary Hesse argues that "almost any model or interpretation carries some surplus meaning. If, however, a model is used in a way that exploits this surplus meaning in prediction and explanation, we shall call it a theoretical model." See Hesse, "Models and Analogy in Science" in Paul Edwards, ed., *The Encyclopedia of Philosophy* (New York: Macmillan and Free Press, 1967), 354-59.

<sup>13</sup> For example, waves provide a model for both water motion and light; similarly, the

pretation of fundamental assumptions and theoretical constructs. This richer interpretive structure (as compared to the tighter correspondences in the model) provides for greater richness of explanation. Through it, the theory maintains a greater open-endedness and a surplus meaning which guides revision and extension of the model. A further source of theoretical richness lies in the multiple models that may be contained within a theory and that emerge according to specific parametric conditions. Through these models, seemingly different phenomena—perhaps varying due to contextual factors—can be understood within the same theoretical framework.

Game theory viewed as a *theory* of intentional behavior illustrates the relation between a theory and its model. The assumption of rationality allows interpretation of solutions in terms of intentional behavior of actors. Specification of different political-institutional environments (e.g., the capitalist market in economic oligopoly theory versus anarchic international society in balance-of-power theory) determines rules of the game that result in different interpretations of models, and ultimately in different models, as the rules are more explicitly introduced into the analysis. Alternatively, different specifications of actors' policy choices and/or preferences may result in different games (e.g., Chicken versus Prisoners' Dilemma) arising within one fundamental theoretical interpretation of the international system. A primary virtue of game theory as theory is the enormous diversity of models contained within it.

For game theory to be a theory of international politics (rather than a general theory of strategic behavior), specific empirical assumptions (correspondences) are required. For example, by assuming that power-maximizing states are the principal actors, game theory subsumes the Realist position. But the game-theoretic approach is not coincident with Realism. While it necessarily treats actors as rational, its empirical assumptions need assert neither that the key actors are states nor that they maximize power. With a different specification of these assumptions, game theory is equally consistent with a modified structural approach.<sup>14</sup>

---

coupled differential equations which can be interpreted as a model of an arms race in international relations may have a very different meaning in thermodynamics. Indeed, it is impossible to speak of a model of something—as opposed to a purely logical and empirically uninterpreted model—without theory to guide us on the correspondence rules. Any empirical model must be embedded in a theory. However, the theoretical richness is often tightly circumscribed in the model. A good example is the use of the “as if” assumption to establish a correspondence at a purely observational level without plumbing the deeper implication of the observed behavior. (See note 17.)

<sup>14</sup> Robert O. Keohane, “Theory of World Politics: Structural Realism and Beyond,” in Ada Finifter, ed., *Political Science: The State of the Discipline* (Washington, DC: American Political Science Association, 1983).

Thus it brings contending perspectives into the same framework and emphasizes their different empirical assumptions.

As a general theoretical approach to international politics, game theory does assume goal-seeking behavior in the absence of centralized, authoritative institutions. It thereby illuminates the fundamental issues of international anarchy and the implications of different configurations of national interests and political circumstances for international conflict and cooperation. At its best, it uses simplifying assumptions to expand our range of understanding and to provide deeper interpretations of international politics.

Metaphor, analogy, model and theory are complementary in social scientific research; they are each appropriate at various stages. It is often useful to go back and forth among them. As research advances, however, metaphor and analogy are of increasingly limited usefulness. The greater rigor and deductive power of the model, together with the interpretive richness and open-endedness of its corresponding theoretical framework, make that combination ultimately more productive.<sup>15</sup>

International political analysis is ripe for a transition to game theory as a theory of international politics. Game metaphors and analogies already are widely used to illuminate and clarify international issues. To apply the deductive power of game theory directly, we must tighten up correspondences between empirical situations and game models, and separate assumptions from predictions. As we do this, the model and theory will provide a guide to relevant empirical evidence just as the evidence will provide a guide for evaluating and revising the model. Treated this way, game theory becomes more than just a new language in which to rewrite history or to restate our arguments. It becomes a powerful tool for expanding our understanding and for stimulating research. The price we pay for such power lies in the assumptions and work required to link the deductive logic to empirical reality. In the remainder of this paper, I address issues raised in using game theory, both as model and as theory, for the empirical investigation of international politics.

## II. USING GAME THEORY TO BUILD THEORY IN INTERNATIONAL RELATIONS

Concepts of game theory provide a guide for constructing theory in international relations. The most fundamental concepts—strategy, stra-

<sup>15</sup> Here we can agree that “perhaps every science must start with metaphor and end with

tegic rationality, preferences, and payoffs—and their implications for understanding international politics will be discussed in this section. Although I will refer mainly to simple  $2 \times 2$  games, the discussion applies equally to other game models. Those will be more explicitly covered in the section on incorporating contextual features of international issues into game models.

#### A. STRATEGIES

A strategy is a complete plan for action, covering all contingencies including random exogenous events as well as endogenous behavior by others. For any but the most trivial decisions, this conception is hopelessly complicated and beyond the calculating power of any man, machine, or state bureaucracy. Although it is possible to treat strategies as simplifications of more complex decision processes, it is more fruitful to treat them as simplified representations of general policy stances. In trade policy, for example, it is meaningful to speak—on a broad level—of strategies of “free” versus “restricted” international trade without worrying about the myriad of nuances such as differential treatment of steel versus textiles, or the use of tariff versus nontariff barriers. Similarly, strategies of reducing military spending, of increasing international tensions, or of promoting environmental protection are each meaningful without supposing that someone has compiled an exact listing of how to pursue that policy under every conceivable contingency. The simple  $2 \times 2$  game pursues this to its logical extreme where only two choices (often too persuasively labeled “cooperate” and “not cooperate”) are available. Although not without limitations, this assumption is a useful simplification for illuminating the fundamental nature of an issue area.

The notion of a strategy is so comprehensive that it can encompass a wide range of phenomena. This breadth of coverage is a central advantage of game theory. However, the limitations of such breadth need to be understood to ensure that the idea of a strategy is used to clarify rather than to confuse. One example is how a strategy incorporates dynamic considerations by allowing for contingent planning through time in response to changing circumstances. Some of the most powerful game-theoretic results emerge when dynamic problems are treated as static choices of strategies which actors will play through time. As we shall find in Section III, however, dynamic considerations are sometimes

---

algebra; and perhaps without the metaphor there would never have been any algebra”—although by the argument of this section, the word “model” should be substituted for “algebra.” See Max Black, *Models and Metaphors* (Ithaca, NY: Cornell University Press, 1962), 242.

obscured when they are subsumed under a static analysis. Even where this is not the case, it will be useful to distinguish simple games that cover a single decision-making period from *sequences* of simple games (i.e., supergames) that represent continued play over a number of periods.

A second concern is over the relationship between having a strategy and behaving rationally. Here the distinction between model and theory is useful. We can agree with Axelrod that, viewed as a model, an "organism does not need a brain to employ a strategy."<sup>16</sup> He demonstrates important results by using game models for a wide range of phenomena where "strategies" have little to do with planning or strategic behavior in the usual sense of the word. For example, game models combined with the assumptions of natural selection offer suggestive explanations for the behavior of bacteria, microbes, and other lower-order life. Similar analyses can be applied to economic and political processes. Thus the models of game theory can be useful for theories not involving rational actors. But, as Axelrod points out, "game-playing becomes richer" in the context of a theoretical understanding where strategies are related to intentional behavior. Explicit use of such strategic rationality in the theory of games captures important aspects of international politics not found in game models divorced from the rationality assumption.<sup>17</sup>

#### B. STRATEGIC RATIONALITY

The cornerstone of Realism is its treatment of states as rational actors. This requires only that states make logically correct calculations in using available information to pursue well-defined goals. This assumption has been subjected to an onslaught of attacks from advocates of theoretical alternatives including bureaucratic politics, psychological models of decision making, social choice, and complex organizations. The best of these critiques provide significant lessons that the game-theoretic tradition needs to address—as will be discussed in the following subsection on the derivation of game payoffs. Nevertheless, even though claims for its exclusive usefulness are surely exaggerated, the value of the rationality

<sup>16</sup> Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984), chap. 5.

<sup>17</sup> This position agrees with Milton Friedman's well-known "as if" argument on one level, but differs from it on another. Friedman's argument is that it does not matter if the actors being modeled actually make (strategic) calculations as long as they act "as if" they did. For him, the proof of the pudding is the accuracy of the predictions that result from the assumption. But if we are to understand and explain behavior in addition to predicting it, his argument will be insufficient. To understand state behavior in international politics, and to avoid *post hoc* reconstruction of behavior as "rational," we must pay attention to the nature and limits of state rationality. See Friedman, "The Methodology of Positive Economics," in Frank Hahn and Martin Hollis, eds., *Philosophy and Economic Theory* (New York: Oxford University Press, 1979), and Mark Blaug, *The Methodology of Economics* (New York: Cambridge University Press, 1980).

assumption for the study of international politics has largely withstood these challenges.

The interdependence perspective, and the game-theoretic tradition more generally, pose a different challenge to the conception of *nonstrategic rationality* that dominates simplistic (but nonetheless distressingly common) views of Realism.<sup>18</sup> Rationality in this Realist world centers on the struggle for power in an anarchic environment. States fend for themselves as they pursue their contradictory interests. Because of the conflictual nature of this “self-help” environment, the situation is mistakenly seen as zero-sum where no cooperation is possible and states can pursue their own best interests without regard for the interests of others. The game-theory perspective reveals that these circumstances prevail only in artificially constructed, two-player parlor games. For real international issues, states’ interests will not be properly characterized by assuming such pure opposition of interests. This leads directly to a *strategic rationality* which incorporates the realization that pursuit of egoistic interest requires consideration of interactions of one state’s choices with other states’ choices. No state can choose its best strategy or attain its best outcome independent of choices made by others. The related substantive implication is that national policy makers need to pursue opportunities for cooperative interactions even as they seek to protect against conflictual interactions.

Two aspects of rationality are especially important for game-theoretic analyses.<sup>19</sup> The first, common to both nonstrategic and strategic conceptions, is the ability to forgo short-run advantages for longer-run considerations. The second, which is the distinguishing trait of strategic rationality, is that actors choose courses of action based on preferences and expectations of how others will behave. Thus, when a state undertakes a certain action, it does not necessarily follow that the immediate result is itself a preferred end for that state. It could be a strategically planned means to some other objective. (An example considered below is the use of Tit-for-Tat strategies to elicit long-run cooperation sometimes at short-run cost.) This possibility requires us to consider how each choice is interrelated with prior and subsequent choices, and to understand national goals independently from observed behavior.

<sup>18</sup> Strains of this narrow interpretation of rationality are apparent even among the best proponents of Realism. See, for example, Waltz (fn. 8), 70. Sophisticated versions of Realism—and certainly those that have incorporated game-theoretic notions—have employed an understanding of strategic rationality. For a clear-headed discussion of narrow rationality, see Bueno de Mesquita (fn. 1), 29–33.

<sup>19</sup> See Jon Elster, *Ulysses and the Sirens: Studies in Rationality and Irrationality* (New York: Cambridge University Press, 1979).



The assumption of strategic rationality is fundamental to a game-theory interpretation of international politics. Individual actions and collective outcomes are understood in terms of states' strategic pursuit of self-interest. The development of institutional arrangements such as regimes is explained in terms of efforts to overcome problems of collective action by altering the "rules of the game."<sup>20</sup> The prospects for further cooperation as well as the dangers of increased conflict can be investigated in terms of the strategic possibilities facing states. As is often the case in theoretical enterprises, the stronger the (rationality) assumptions made, the richer the interpretation provided by game theory. Conversely, if the rationality assumption is seriously circumscribed by (say) bureaucratic, psychological, or organizational factors, the same models need to be interpreted differently.

Finally, strategic rationality lies at the heart of one of the most attractive features of game theory. By placing the rational choice of state policy at the forefront of the explanation, game theory allows for an autonomy in state choice even as it predicts and explains those choices deterministically through an understanding of the overall strategic interaction. Thus the game model combines purposeful behavior with a specification of the structure of international politics which constrains that behavior. It links systematic macro-theory to voluntaristic decisions. States have choices; but the choices they make are determined to a greater or lesser extent by the exigencies of international politics.<sup>21</sup>

### C. PAYOFFS

Rationality assumes that states pursue goals, but those goals are not specified. A game-theoretic perspective requires analyzing states' motivations and how their preferences map into payoffs within a game model. Establishing this correspondence between an issue area and its game model is the toughest problem confronting successful empirical application of game theory. To do so, we must posit national goals that depend on internal values as well as on external circumstances, thereby incorporating other empirical and theoretical understandings of international issues into the model.

For an inductive derivation of payoffs, *what* has happened cannot serve as an explanation of *why* it happened. Simple-minded uses of the "revealed preference" approach, which Sen aptly labels a "robust piece of *evasion*," lead to circular reasoning from the choices made by actors

<sup>20</sup> See Keohane (fn. 6), and Snidal (fn. 6).

<sup>21</sup> Pierre Allan, *Crisis Bargaining and the Arms Race* (Cambridge, MA: Ballinger, 1983), 5-6, and Hahn and Hollis (fn. 17), 15.

back to their preferences.<sup>22</sup> Except for metaphorical and purely descriptive purposes, nothing is gained by this sort of restatement of the situation. To avoid circularity, revealed preference can be applied in a somewhat different way. Choices made by a state in other situations where it faced similar circumstances can be used to infer preferences (and hence payoffs) inductively. Now the evidence used to determine the game payoff structure is independent of the outcome. Further independent evidence about preferences may be available in archives and memoirs, including statements of leaders about their objectives. Such sources require a careful and systematic approach to the collection of evidence; self-serving statements of leaders need to be ruled out, and biased or *ad hoc* collections of anecdotes in support of a postulated payoff structure must be avoided.

But even the most rigorous inductive approaches are problematic for game models because strategic rationality casts doubt on any such interpretation of relevant evidence. Since actors sometimes forgo immediate interest for longer-term gain, observed action may not reflect preferences directly. Furthermore, the essence of strategic behavior is that a state may forgo individually optimal actions to collaborate with others in achieving mutually preferred outcomes (e.g., the observation of “co-operative” behavior in Prisoners’ Dilemma should not lead us to the incorrect inference that a state’s highest individual payoffs are associated with such behavior). Observed behavior of strategic actors is thus often a biased and unsatisfactory indicator of underlying interests.

Finally, inductive procedures provide only an incomplete map of preferences. With regard to revealed preference, only choices actually made by a state are directly observed; evidence about their evaluation of other possible outcomes is indirect and incomplete. In terms of simple games, direct evidence is available on only one of the four cells. This paucity of information is even more striking because the 2 x 2 game is already a simplified representation of a much richer set of preference mappings. Thus, even if augmented by other sources, purely inductive evidence will generally be inadequate for determining the structure of a game matrix.

Theoretical specification of states’ preferences can provide a route out of this empirical quandary. Such “theories of the payoffs” necessarily precede the game model. They enable game theory to be constructively and complementarily linked to other approaches to international politics—even, in some cases, to theories that may be viewed as alternatives

<sup>22</sup> Armatya Sen, “Rational Fools: A Critique of the Behavioral Foundations of Economic Theory,” in Hahn and Hollis (fn. 17), 92; emphasis in original.

to rational models. The possibilities for developing a theoretical basis for the derivation of payoffs are best seen by considering alternative ways to accomplish this goal.

When theoretical knowledge of an issue area is substantial, assessing states' preferences is straightforward. This knowledge provides a way to model a situation and derive its payoff structure. Consider the neo-classical theory of international trade. It provides an elaborate framework for deriving the payoffs determined by strategy choices which might include the imposition of tariffs, restrictions on capital flows, or subsidization of industries. Such strong theory allows precise specification of the strategic situation between states over a host of important trade questions. A second example is found in the literature on military spending. Theoretical arguments about the impact of military expenditures among allies and rivals, combined with knowledge of prevailing international tensions and alliances, can reveal the underlying strategic structure of these situations.<sup>23</sup>

But states' preferences may not always be tightly linked to objective understandings of an issue area. Perceptions and information processing, as well as organizational or bureaucratic imperatives, may change the relevant payoffs for decision makers. Theoretical understanding of such factors may illuminate additional considerations that influence states' decisions on foreign policy alternatives. Of course, extreme versions which explain behavior largely in terms of the decision process itself will compete with, rather than complement, game models. But less extreme versions which leave a role for intentional behavior will suggest relevant empirical factors that affect strategic behavior. For example, Keohane incorporates "bounded rationality" into decisions to participate in international regimes, while Allan uses the concept of "diplomatic climate" to admit cognitive considerations of past events into rational models. When rational action is circumscribed but not overwhelmed by such factors, theoretical integration of this sort will be productive.<sup>24</sup>

Theoretical considerations are also valuable when partial information restricts the logical possibilities for the payoff structure. For example, if a situation is symmetric or if the nature of common and conflicting interests is understood, then the set of possible 2 x 2 games can be considerably reduced.<sup>25</sup> Alternatively, substantive knowledge that an

<sup>23</sup> Mancur Olson and Richard Zeckhauser, "An Economic Theory of Alliances," *Review of Economics and Statistics* 48 (No. 3, 1966), 226-79; Snidal (fn. 6), chaps. 5-6.

<sup>24</sup> Keohane (fn. 6), chap. 7, and Allan (fn. 21), chap. 4. There are limits to rational theory, however: it cannot always incorporate contending approaches except in a trivializing way. Even some issues that are internal to the theory—especially problems of preference aggregation in determining a "national interest"—are far from resolved.

<sup>25</sup> Russell Hardin demonstrates how to narrow the range of relevant games for deterrence

issue area is characterized by relative rather than absolute gains (for instance, mercantilist versus liberal trade policies), or about the way policy choices are interrelated across states (for instance, international transmission of macroeconomic policies among economically integrated states) will help to describe the game structure. Information about the general political-institutional setting for relations among states, about the likelihood of an issue continuing into the future, or about the number and relative size of the states involved—all of which are discussed separately below—will also help to identify the structure of strategic interrelationships.

Theoretically inspired approaches to inferring payoff structures have several advantages over purely inductive approaches. First, by focusing on underlying motivations rather than on observed outcomes, preferences are distinguished from actions, and individual interests from strategic calculations. Theoretical approaches also provide more complete information on the payoffs for outcomes beyond the one that actually occurs (e.g., on all of the “cells” of the  $2 \times 2$  game), which permits an analysis of the complete game structure. Finally, by forcing us to investigate new questions concerning the motivations underlying the behavior of states, they ensure that the analysis is not simply a redescription of the issue in more formal language.

A second advantage of theoretical derivation is that it systematically addresses the question of changing preferences through time and/or changes in the institutional environment of international politics. One requirement of the game-theoretic approach is that actors have reasonably stable preferences and behave consistently when confronted with comparable choices. If preferences change too quickly, the model degenerates to a generalized *post hoc* revealed-preference exercise, where actions are assumed to reflect prevailing fluctuations in preferences. A theoretical approach avoids this circular reasoning by incorporating those factors that affect preferences and payoffs through time (including prior outcomes) directly into the theory of payoffs. Such systematic treatment of changing preferences or evolving institutions allows for a properly dynamic treatment of international issues.

But the greatest advantage of theoretical specification of payoffs is that it unleashes the deductive power of game theory. By combining the game structure defined by preferences and available strategies with game theory solution concepts, we are led to new inferences about the behavior of individual states and about the overall outcome. These (falsifiable)

---

problems in “Unilateral Versus Mutual Disarmament,” *Philosophy and Public Affairs* 12 (Summer 1983), 236–54.

predictions are independent of observed behavior, which provides evidence on the correctness of the assumptions and the usefulness of the game-theoretic approach. If the theory withstands the test, these new deductions expand our knowledge and understanding of international behavior, and the analysis moves beyond metaphor and analogy to models and theory.<sup>26</sup>

A final implication of theoretical derivations for payoffs is that any particular application is not a test of the game-theoretic approach in isolation, but of game theory in combination with the particular theoretical assumptions embodied in the game structure (for instance, the underlying motivations of actors from which the payoffs are derived). If a particular model is inadequate, the theory of the payoffs will be rejected before the game-theoretic approach itself. The general usefulness of the latter will be rejected only if a more plausible payoff theory cannot be constructed. Even then, we will not be sure whether this failing is really due to our own (possibly temporary) shortcomings in understanding and specifying actor motivations. Nevertheless, the theoretical specification of payoffs makes the game model more vulnerable to empirical evidence and leaves it potentially falsifiable. This condition qualifies it as a serious explanation and more than just a tautological redescription of the world. Finally, attempts to revise the theoretical specification in order to accommodate contradictory evidence are likely to lead to innovations in our understanding of international politics.

### III. THE CONTEXT OF INTERNATIONAL POLITICS AND GAME MODELS

One common but misguided criticism of game models is that they are too "simple" to capture the complexity of international politics. Game models may not be able to capture all details of international interactions, but in earlier sections of this paper I have argued that simplicity actually enhances the power of the theory for grasping complexity. I will now expand that argument by looking at extensions of simple game models to capture key contextual factors of international politics. The theoretical use of game models allows us to adapt them directly to the most salient aspects of international politics.

<sup>26</sup> Examples of these deductions include those discussed above. The success of rational-actor approaches in other areas of political science is due to precisely this sort of approach (for example, the assumption that candidates maximize votes leads to conclusions about their behavior). For a discussion of this (and a critique of the trivializing use of revealed preferences by imputing utility to "citizen's duty" to explain voting), see Barry (fn. 9), chap. 2.

## A. EXTERNAL GAME ENVIRONMENT AND INTERNATIONAL REGIMES

The goal in specifying a game structure is to capture the essential features of an international situation. Some of these are assumptions (actors, strategies, payoffs) while others are predictions (the outcome). The value of a game representation depends on how successfully it captures significant aspects of the international environment. Game theory also allows for broader interpretations, so that features not explicitly entered into a model may be useful for understanding and interpreting it.

The specification of actors, capabilities, and preferences defines an overall game model of a specific issue. Each state's strategic interconnection with others outlines "structural" constraints that determine its opportunities. This notion of structure is atomistic in viewing structure as simply the sum of the individual units and emphasizes the structure of individual issues rather than that of the overall international system. It is therefore inadequate for examining certain conceptions of world structures.<sup>27</sup> Nevertheless, the international system, with its established patterns of practice and rules, is significant for defining the individual game model and for deriving conclusions from it. In this way, the structure of international politics modeled as the rules of the game is distinguished from the behavior of states within those rules. (Of course, the rules of the game may be altered through establishment of new patterns of behavior, expectations, and norms, which emerge as new rules in the longer term.)

Indeed, even the definition of issues, actors, and choices depends on the preexisting international order. For example, separation of trading issues into the GATT framework, recognition of states as relevant decision makers, and acceptance of restraint from force in deciding trade disputes all reflect theoretical assumptions as to which aspects of the existing trade *regime* can be taken as "given" for analysis of the politics of international trade. The assumptions of one analysis may be the objects of investigation in another. Regimes are not independent of the existing global political structure, but are built upon it and "nested" so that the fabric of one provides the foundation of another. Therefore the rules of the existing international order define the underlying game, even as the game is used to pose further questions about the development of international regimes.

Empirical issues concerning the emergence of regimes are related to

<sup>27</sup> Richard Ashley, "The Poverty of Neo-Realism," *International Organization* 38 (Spring 1984), 236-86.

theoretical issues raised by solutions within game theory. For individual states, these issues include the concept of *maximizing self-interest* (e.g., maximax versus maximin strategy choices), and how this is affected by the norms, conventions, and expectations that emerge in the regime. This leads to the question of *efficiency*, or whether the regime enables states to exploit mutual interests. A third question is whether the outcome is *self-enforcing* (i.e., a stable equilibrium) so that the regime will be maintained by the self-interested behavior of states without centralized enforcement. A final concern is the distribution of benefits and considerations of *fairness* which affect our evaluation of a regime.

These central questions for investigating the role of international institutions in promoting international cooperation provide criteria for predicting and evaluating behavior within the regime as well as transformation of the regime itself. This fuller interpretation of the game model is in terms of the theoretical assumptions of strategic rationality and interdependent decision making that are fundamental to game theory.

#### B. PAYOFF MEASUREMENT FOR ANALYZING BARGAINING AND ASYMMETRY

Stronger assumptions about preferences—for example, interval versus ordinal measurement—contain more information for deriving stronger conclusions from game models. The appropriate level of measurement depends on our knowledge of the issue as well as the particular theoretical questions to be investigated. Ordinal measurement is typically the minimal level that provides a sufficiently clear definition of the game structure for fruitful game analysis, but it is not always adequate.<sup>28</sup> Stronger measurement is necessary for analysis of certain central questions. Consider some key examples: the result that repetition of an issue facilitates cooperation is based on interval-level comparison of alternative streams of (discounted) payoffs through time. Cooperation through issue linkage, viewed in terms of “trading” assistance on one issue for assistance on another, depends on interval-level weighing of benefits gained in one issue against costs incurred in the other. Indeed, solutions to bilateral bargaining problems can be shown, in general, to depend on interval-level payoffs. Even a simple  $2 \times 2$  game can be dramatically

<sup>28</sup> Ordinal payoffs correspond to “first,” “second,” and so forth. Interval measurement requires meaningful “units” (e.g., degrees of temperature or units of payoff) for the “distance” between outcomes (e.g., change in the temperature or in a state’s payoff). Cardinal measurement requires a meaningful “zero” (e.g., absolute zero in temperature scales) and is largely irrelevant for game theory. Other levels of measurement may fall between these categories (e.g., partial orderings may give us interval-level comparisons between some outcomes, but no direct comparison between others).

altered by replacing its ordinal payoffs with interval-level payoffs that are consistent with the original ordinal ranking. In brief, interval-level payoffs will be necessary, in addition to the simple order of preferences, whenever intensity matters.

Where interval-level payoffs are not available, partial information about intensity may still be useful. For example, the incentive to defect on a cooperative agreement might be known to be greater in one situation than in another, even if the comparison cannot be made more precise. The greater incentives will tend to make decentralized agreements less effective and decrease the prospects for cooperation. Thus, ordinal payoffs can often be augmented by partial information on preference intensity.<sup>29</sup>

Intensities of preferences differ not only *for* actors but *across* actors. The latter situation raises the thorny issue of “interpersonal” comparisons of utility, whereby relative welfare gains of states are compared. Such considerations are central in normative evaluation of outcomes (for instance, the distributive implications of alternative Law of the Sea proposals), and may thereby influence actions of states (such as granting tariff preferences to less developed countries). Interpersonal comparisons are also important in purely “positive” analysis, however. The ability to do great harm (or confer benefit) to another at a relatively low cost may affect a state’s behavior. This is especially true for situations of asymmetry where one state is substantially more vulnerable than another. Less vulnerable states will use their position advantageously to determine the outcome; when both are equally vulnerable, outcomes are more likely to be equal.<sup>30</sup> It is treacherous but essential to make such comparisons—especially where issues are extremely asymmetric. Ordinal game models often make issues appear symmetric (for instance, by treating a very large state and a very small one as equal partners in a Prisoners’ Dilemma with the “same” ordinal preferences) even though they are extremely asymmetric under interpersonal comparison and interval measurement (if, say, the issue is vital to the small country but inconsequential to the large one). Since many key problems of international politics revolve around questions of interdependence versus dependence, this consideration is important.

<sup>29</sup> For example, see Robert Jervis, “Cooperation under the Security Dilemma,” *World Politics* 30 (January 1978), 167–214, at 174. Nevertheless, ordinal payoffs can carry our analysis very far, and economists once (erroneously) even believed they were sufficient for virtually all purposes. An interesting account of how measurement is integrally related to the questions we are investigating is Robert Cooter and Peter Rappoport, “Were the Ordinalists Wrong About Welfare Economics?,” *Journal of Economic Literature* 22 (June 1984), 507–30. For a thoughtful but more technical discussion of these issues, see Shubik (fn. 4), chaps. 4 and 5.

<sup>30</sup> For examples of the impact of vulnerability, see Jervis (fn. 29), 171–73.



Thus, problems of bargaining (including linkage and cooperation through time) and of asymmetry can be captured in game-theoretic models that include stronger theoretical attention to underlying preferences and stronger measurement of payoffs. Such game models also provide a handle on important normative questions surrounding positive analysis. While these various problems are difficult to deal with, they are not intrinsic to game theory analysis; the latter only serves to illuminate some of the substantive and theoretical shortcomings in our understanding of international politics.

#### C. ITERATED AND DYNAMIC GAMES: TWO ACTORS PLAYING THROUGH TIME

International politics is inherently dynamic and involves interactions among states through time. As discussed above, a strategy can encompass dynamic situations within a purely static analysis (that is, where states choose entire future courses of action at once). But compressing all of these possibilities into a single choice may involve unrealistic assumptions about the capacity of states as decision makers, and obscure important aspects of international politics. To understand the impact of evolving international regimes, of changing expectations, of learning and adjustment by states, or of changing national preferences, the concept of "change through time" needs to be clarified.<sup>31</sup>

The term "dynamic" is not as precise as its mathematical connotation implies, and it has many different usages. For international politics it emphasizes (1) the impact of states making multiple decisions through time, and (2) mutual adjustment among states through time. It is of particular importance whether new equilibrium outcomes emerge when the game is played through time, and whether these are stable. These questions (rather than the form of the game) distinguish dynamic analysis.

Treating international issues as dynamic is complicated by the problem of defining the decision-making period. Discrete time models have predominated because they are easier to work with. The standard treatment has been in terms of sequential plays of iterated games where time is divided into decision periods corresponding to a single play of a static game. However, changes in "revealed" preferences or contextual factors between iterations should not be invoked on an *ad hoc* basis to define different game iterations or to fit different game models to an issue at different points in time. Instead, dynamic analysis requires theoretically informed understandings of time-changing empirical patterns. For ex-

<sup>31</sup> Complications of inconsistent time preferences are ignored here. See Elster (fn. 19), chap. 2.

ample, superpower "succession crises" due to machinations within the Soviet Politburo or the American electoral process may explain systematic fluctuations in the politics between them. Alternatively, past interactions may affect present behavior, perhaps by increasing interdependence or by changing expectations and trust among nations. Incorporating such effects allows us to adapt a systematic sequence of static models to the changing circumstances of a dynamic world, but it requires stronger theoretical assumptions about international politics (for instance, of exogenous factors that affect the game structure, or endogenous changes in expectations) than those necessary to construct the individual static game models.

A further application of dynamic analysis arises when current behavior is affected by prospects for future play of the game. Continuation of an issue through time may affect a state's behavior in each "play" of the game and alter the (equilibrium) outcome. The power of this analysis is illustrated by its most exciting result to date: whereas cooperation is not individually rational in single-play Prisoners' Dilemma, it will be rational under certain (specified) conditions in the iterated Prisoners' Dilemma supergame.<sup>32</sup> The reason is that states will forgo short-run incentives to defect when they can thereby achieve longer-run benefits from cooperation through time. This makes international cooperation possible in the absence of centralized enforcement and indicates the importance of dynamic analysis of international politics. It is also a prime example of using the deductive power of game theory to derive new results.

Iteration changes not only the outcome of the game, but the underlying structure of the situation. Whereas in single-play Prisoners' Dilemma there is a dominant strategy to defect, and the actions of others are barely relevant, coordination with other states becomes important in the iterated game. For example, the Tit-for-Tat strategy (i.e., reciprocate the other's last move) is effective only when the other state (in a bilateral situation) or a sufficient number of states (in a multilateral world) have adopted a compatible strategy. Thus the fundamental problem facing the individual state is altered once the Prisoners' Dilemma game is iterated. However, cooperation is not always enhanced by iterated plays of other games. For example, prospects of continued plays of coordi-

<sup>32</sup> Cooperation will make sense in anticipation of, and in response to, cooperation by the other party. An early formulation of the impact of iterated Prisoners' Dilemma games is Martin Shubik, "Game Theory and the Paradox of the Prisoner's Dilemma," *Journal of Conflict Resolution* 14 (June 1970), 181-93. The result is worked out formally in Michael Taylor, *Anarchy and Cooperation* (New York: Wiley, 1976) and extended via tournament techniques in Axelrod (fn. 16).

nation games provide incentives to break existing cooperation agreements in order to press for an alternative coordination point that the players prefer. Thus, while Prisoners' Dilemma and the failure of cooperation is not always the proper metaphor for international relations, neither is iterated Prisoners' Dilemma and success in cooperation the proper metaphor. Individual issues need to be modeled separately to understand the impact of iteration.<sup>33</sup>

It is interesting to reflect upon the Prisoners' Dilemma supergame as a "dynamic" game. Since it involves choosing a single strategy for playing the game through time (e.g., play Tit-for-Tat or play Always Defect), it has the formal properties of a static game. It is dynamic only in the sense that strategies are consciously directed toward the problem of how to play in an iterated series of static games. Through this, it provides a systematic analysis of shifts within strategies during a sequence of iterated games (for example, a shift from "cooperate" to "not cooperate" within a Tit-for-Tat strategy). This interpretation of the supergame (plus the fact that it may involve a transformation of the static game as discussed above) gives it dynamic qualities.<sup>34</sup>

The greater power of the supergame requires stronger assumptions that states compare current benefits (for instance, from not cooperating on the current play) with future benefits (for instance, maintaining cooperation through time). Such comparisons necessarily involve interval-level payoffs and more information about preferences (intensities as well as ordering) than do ordinal payoffs. Also, because payoffs from individual iterations are received at different points in time, states must compare them through "discount rates" showing their time preference for immediate versus future benefits. These discount rates depend on domestic circumstances, the situation of the state in the international

<sup>33</sup> The coordination aspects of the two-person Prisoners' Dilemma supergame can be seen in the matrix of supergame strategies in Taylor (fn. 32), 39. For a discussion of the differences between iterated Prisoners' Dilemma and iterated Coordination, see Duncan Snidal, "Coordination Versus Prisoners' Dilemma: Implications for International Cooperation and Regimes," *American Political Science Review* (forthcoming, December 1985).

<sup>34</sup> A further dynamic adjustment process in terms of the evolution of strategies, whereby more successful strategies in one period are more likely (for reasons of survival, imitation, or learning) to occur in subsequent periods, is added by Axelrod (fn. 16). R. Harrison Wagner, "Theory of Games and International Cooperation," *American Political Science Review* 77 (June 1983), 330-46, provides a useful critique of the attempt to embody dynamic assumptions in a single 2 x 2 game through the use of sequential games that deny "players the opportunity to cheat (by assuming that they will cooperate conditionally) . . ." (pp. 332-33). However, he commits a similar error in assuming conditional behavior within the extensive game (p. 344). Iterated game analysis keeps the decision period much cleaner and less subject to artificial insertion of "conditional" cooperation that is based, in effect, on the ability either to predict the future or to recover from an adversary's behavior before payoffs accrue.

system, and uncertainty about continuation of game iteration into the future.<sup>35</sup>

One important factor is the number of iterations in a game. The longer an issue is expected to persist, the greater the impact that future play will have on current choices. Similarly, the more quickly states can adjust their policies, the shorter the time frame of any single game, and hence the more iterations in any fixed period of time. This suggests that states may promote cooperation (in Prisoners' Dilemma) by improving verification in arms control, or by dividing negotiations into larger numbers of smaller sequential steps. These policies will be enhanced by perceptions of other states as reliable partners for future cooperation, but diminished by incentives to take a "final move" and sacrifice future collaboration for immediate advantage.

In addition to providing an insight into incentives to cooperate, iterated game analysis raises questions of how states adjust to new equilibrium outcomes. Axelrod's work explores the evolution of cooperation through the natural superiority of certain strategies, especially Tit-for-Tat, in iterated games. However, while the evolutionary *model* is insightful for understanding the emergence of cooperation, evolutionary *theory* is misleading in its specification of the mechanism of adjustment. This is because the international system has neither the selective elimination (states are rarely eliminated even in war) nor the random variation that evolutionary theory requires. Instead, there is rational adjustment through learning and planning of a sort that is better captured by game theory. Even though its model of behavior is identical (for example, in showing the efficacy of Tit-for-Tat), rational adaptation allows for a more compelling and fruitful theoretical interpretation of international politics and cooperation than does evolutionary theory.<sup>36</sup>

Interpretation of dynamic games in terms of strategic rationality directs research to the incentives underlying the emergence of international regimes that facilitate cooperation. Interactions are understood in terms of the deliberate behavior of states trying to improve their welfare

<sup>35</sup> The formal results for cooperation through time require either that the game continue forever or that there be uncertainty about its termination date. Luce and Raiffa (fn. 4) show that cooperation will not be rational if the termination date is known. Russell Hardin argues in *Collective Action* (Baltimore: Johns Hopkins University Press, 1982) that since this is implausible, future play will provide incentives to cooperate.

<sup>36</sup> This is a case where the same model is a model for two different theories (e.g., evolutionary and rational). Axelrod's discussion (fn. 16) recognizes the alternative interpretation of his model in terms of learning and adaptation, especially in his chapter on cooperation in trench warfare. Nowhere does he provide a vulgar evolutionary view of politics. Nevertheless, it is important to emphasize the very different interpretations of his model of iterated play under rational as opposed to evolutionary theory. See Elster (fn. 19), chap. 1, and Keohane (fn. 14).

through cooperation. Acceptance of the rules and norms of regimes is interpreted in terms of maximizing behavior: regimes are effective and persist only as long as states find it in their interest to maintain them. This rational adjustment mechanism explains cooperative behavior even when there is no threat of elimination and when the environment of international politics is changing too rapidly for evolutionary adjustment to ensure successful adaptation. It also explains why cooperation might emerge on global issues where all nations are affected equally, so that evolutionary selection among them according to their behavior is impossible. At the very least, we should hope that adjustment to cooperation will be rational in a world of nuclear weapons.

#### D. A MULTILATERAL WORLD: N-PERSON GAMES

Game models with more than two actors are important since many international issues involve larger numbers of states (for instance, five actors in the classic European balance of power, many participants in the Law of the Sea, and the "Groups" of five, ten, seventy-seven, etc., in contemporary economic regimes). A consideration of N-person game-theory models suggests how such extensions can be helpful for some problems, but are of only limited usefulness for others.<sup>37</sup>

N-person games are complicated and provide few simple "answers," but a number of interrelated assumptions can simplify the analysis of specific situations. The first simplification pertains when the *number of states is large*, so that no state has a substantial impact on any other state. Each can assume that its actions are unnoticed and pursue its interests on the assumption that others will not react. This reduces the problem to narrow rationality—equivalent to the one-player "game against nature"—and poses no problems of strategic complexity. That is not to say that such situations are without interest or are nonproblematic: collective-action problems can lead to collectively suboptimal outcomes.

<sup>37</sup> When should a situation be treated as an N-person game? In some cases the answer is obvious because it is technically impossible for the actions of many states to be insulated from one another. For example, in conservation of fish stocks in a "commons," all states fishing that commons will be relevant. The "N-ness" of the problem will depend on the exact nature of the commons. If fish species are nonmigratory and the commons is territorially divided, then only territorial states (perhaps as few as one) need be involved. But if species are migratory and/or the commons is not divided, then any state that fishes those species may be a relevant actor. This determination may be complicated if states act strategically and misrepresent the extent of their interest in the issue, or if the number of participants itself is not exogenous to the regime. For example, in the construction of economic regimes (e.g., trading blocs), deciding the scope of membership may hinge on the expected impact on the regime. An overview of the technical game theory material is available in Anatol Rapoport, *N-Person Game Theory* (Ann Arbor: University of Michigan Press, 1970) and in Shubik (fn. 4).

The large-number assumption, however, is not always fruitful for international politics because issues often involve intermediate numbers of states with the capacity for monitoring and reacting to each others' actions—especially if there is the possibility of discrimination as discussed below.

A related simplification is the assumption of *symmetry*. This pertains when states face similar opportunities, have similar interests, and are of approximately equal impact in an issue area. In such cases, the relation of each state to every other is roughly the same; multilateral relations can be understood in terms of the strategic relation of any one state to all of the others. Schelling's and Hardin's analyses of "k-groups," and many problems of public-good provision are examples.<sup>38</sup> These symmetrical models can sometimes even be modified to incorporate certain asymmetries within them (for instance, the introduction of different distributions of interest into a symmetric problem of public-good provision).<sup>39</sup>

A final simplification concerns the ability of states to *discriminate* their actions with respect to other states. For example, states differentiate their tariff schedules (e.g., common markets or the Generalized System of Preferences) and use military force to defend allies while threatening enemies. To be effective, discrimination must be linked to strategy choices of other states so that it can be made contingent on their behavior. In this respect, perfect discrimination is the opposite situation from that of large numbers. When discrimination is perfect, each state can adopt a separate policy toward every other state, and the N-person game can be analyzed as a set of *linked* two-person games. Axelrod's analysis of the evolution of cooperation relies on this assumption since actors are assumed to be capable of cooperating with some actors while not cooperating with others. By using the discrimination assumption, he derives a compelling account of the emergence of cooperation among "N" actors while relying largely on two-actor game theory.<sup>40</sup>

Discrimination is not always a reasonable assumption. It may be difficult on technical grounds having to do with the nature of the issue or the capacities of states. For example, problems in the provision of

<sup>38</sup> Hardin (fn. 35); Thomas Schelling, *Micromotives and Macrobehavior* (New York: Norton, 1978).

<sup>39</sup> Snidal (fn. 3).

<sup>40</sup> Axelrod's analysis has every actor playing the same strategy (e.g., Tit-for-Tat) against every other actor; but that may mean behaving differently vis-à-vis different states on any particular move (according to how they behaved on the previous turn). However, the linked nature of the 2 x 2 games is central to his analysis since the evolutionary survival of actors depends on comparisons of how each fares (on average) against all the others. Axelrod (fn. 16), chap. 3.

international public goods arise precisely because states cannot isolate their actions toward one state from those toward another (e.g., restraint in the size of fishing catches from international waters). Even if feasible, discrimination may be hard to enforce. Thus, discriminatory trading arrangements are susceptible to evasion through transshipping, and it is usually difficult to threaten one state militarily without making others nervous. Even where discrimination is technically feasible, political arrangements such as the norm of nondiscrimination in trade may make it impossible. In other cases, political institutions such as alliances or trading communities are designed to facilitate discrimination. When it pertains, discrimination allows analysis of N-person games in terms of simpler models.

Many N-person situations will not fall under the related assumptions of large numbers, symmetry, or discrimination.<sup>41</sup> What can be done to analyze these important situations in international politics? First, there are a few general conclusions from N-person game theory that can increase our understanding of international politics. Although N-person game theory often disappoints by providing a multiplicity of solutions instead of a single one to any particular game, this multiplicity itself provides insight into the nature of strategic problems.<sup>42</sup> Differences among solutions highlight important aspects of the strategic structure for understanding politics among nations. For example, conflict between the criteria of efficiency (Pareto-optimality) and of maximizing national self-interest highlights the inherent conflict between collective and individual rationality in the problem of collective action. Similarly, analysis of the existence of the "core" solution may indicate problems for co-operation that are likely to arise whenever there are multiple strong and overlapping subcoalitions in a population of states.

Second, more specific applications of N-person game results may be possible through modeling specific international issues. The oligopoly literature has already provided inspiration for international political research; its rich body of theoretical and empirical knowledge undoubtedly contains further insights. However, this knowledge should not be incorporated by analogy (as in Waltz), but by its example of how to

<sup>41</sup> The hardest issues to analyze will be nonsymmetric ones involving intermediate numbers of states with limited capacities to discriminate their actions. That category, of course, covers much of the ground in international politics.

<sup>42</sup> This problem is not unique to N-person games; it also crops up in two-person games (e.g., in the differences between outcomes predicted between maximax versus minimax strategies). While solution concepts sometimes converge in N-person games, they often do not, and the complexity of the strategic structure makes it harder to compare or choose among them than in two-person games. See Rapoport (fn. 37), and Shubik (fn. 4).

construct models and theory. Tailoring models directly to international politics will capture the richness of international politics better than even the most heroic attempts to patch up the analogy between firms and states. For example, the literature on balance of power contains tremendous insights on strategic interrelations among different numbers of states—but much of this work remains inchoate, expressed in terms of descriptive rules of behavior instead of a more formally articulated model in terms of the goals of states. Other work that takes the important step of attempting to model the goals of states directly fails to take into account the strategic interrelationship among them.<sup>43</sup> Integration of these two approaches holds great potential for a more complete theoretical understanding of political-military relations.

Nevertheless, there are few easy results from N-person game theory, and it is far from a panacea. In general, the theory will be inadequate if our goal is to recreate specific situations in our models. This will be especially true for the more complicated N-person dynamic games that have not been discussed here. But insofar as our goal is understanding general problems of international cooperation as the number of states increases, the simplifying assumptions and general conclusions of game theory will be valuable.

#### IV. CONCLUSION: GAME THEORY AND INTERNATIONAL COOPERATION

The ultimate criterion for evaluating game theory is whether it expands our understanding of substantive issues such as those analyzed in the case studies of international cooperation in the essays that follow. The theory is very general and does not provide specific predictions without additional, auxiliary assumptions. Its usefulness therefore depends on whether it poses interesting questions about the politics of international issues and suggests fruitful directions for empirical elaboration, rather than on whether it provides correct answers in any narrow sense. The case studies provide a test of the explanatory power of this empirical elaboration of the theory. Of course, the test is not of specific predictions of behavior in particular circumstances—theories don't make those sorts of predictions—but of general explanations of behavior in issue areas. If explanations in particular case studies are not empirically compelling, it may be either because the general theoretical approach is

<sup>43</sup> Two relevant works to build upon are Morton Kaplan, *System and Process in International Politics* (New York: Wiley, 1957) and Bueno de Mesquita (fn. 1).



deficient or because there are problems in the particular empirical elaboration. Thus the test of the theory is ultimately whether the case studies, taken as a whole, demonstrate that a game theory approach provides new insights to international politics.

Game theory as a theory of international politics has a richness that makes it amenable to such broad empirical application. The theoretical constructs are flexible and can be adapted to different substantive problems. The tremendous variety of models contained within the theory allows for a systematic incorporation of the most salient contextual features pertaining to different issues. In the simple  $2 \times 2$  game, this is reflected in the different game structures (for example, Prisoners' Dilemma versus Chicken) that can be distinguished through the careful modeling of payoffs and of the international environment. Extensions to N-person and dynamic games allow other important dimensions to be addressed. Diverse international issues can be handled within a common game-theoretical framework which does not suppress that diversity, but builds upon it to explore the implications of various contextual differences. In this way, the theory emphasizes the importance of context without becoming lost in it.

This versatility of game models can be a vice rather than a virtue when used improperly. Metaphorical and analogical approaches are valuable for descriptive and expository purposes, but are too flexible to provide falsifiable propositions. The same is true of models applied in a *post hoc* fashion. Only models embedded in theoretical arguments, and carefully tailored to the relevant empirical correspondences (e.g., for payoffs or number of iterations) in an issue area will provide interesting and (potentially) falsifiable empirical claims. Moreover, the deductive power of these models will help to keep inferences separate from assumptions even as they produce new predictions.

It is as a theory of state behavior that games hold their greatest promise for understanding international relations. As a general theory, game theory brings the contending "interdependence" and "Realist" positions together in a common framework. While limited by the correctness of its assumptions about strategic rationality, game theory is sufficiently flexible to incorporate many differing assumptions about world politics and individual issues. Self-interested behavior of states is not prejudged as necessarily leading to either cooperation or conflict. Nor is the emergence of stronger international institutions, formal or informal, seen as either inevitable or utopian. Instead, these conditions are simply among the diverse possibilities that might be predicted depending on the con-

crete circumstances of different issues. The theory pushes us to fill in the appropriate empirical correspondence for investigating such possibilities. In this way, a “game *theory* of international politics” will help us to elucidate and renovate the broader metaphors of “Hobbesian anarchy” and “international organization” that have divided and obscured our understanding of international politics.