REALISM, GAME THEORY, AND COOPERATION

By ROBERT JERVIS *

EVER since Thucydides, scholars have emphasized that international politics is shaped by the anarchical context in which it takes place. The pernicious effect of what John Herz called the security dilemma—the fact that most of the ways in which a country seeks to increase its security have the unintended effect of decreasing the security of others—also is familiar, and indeed, also can be found in Thucydides: "What made war inevitable was the growth of Athenian power and the fear which this caused in Sparta." In view of such dynamics, how can states cooperate? Recent analyses have formalized these problems and analyzed them by means of modern social science techniques.[1]

The work arises from the intersection of Realism and game theory. Although common interests are stressed more than in some forms of Realism, the basic assumptions clearly fit within this school: the focus is on the state as an actor[2] and on the strategies that can rationally be used to further its interests. The studies use simple game-theory models, or what Barry O'Neill calls "proto-game theory,"[3] in order to gain the insights

* I am grateful for comments by Robert Art, John Conybeare, Jeffrey Frieden, Joanne Gowa, Joseph Grieco, Ernst Haas, Stanley Hoffmann, Stephen Krasner, Deborah Larson, Fred Lawson, Helen Milner, Richard Nelson, and Jack Snyder.

[1] John Hertz, "Idealist Internationalism and the Security Dilemma," *World Politics* 2 (January 1950), 157-80; Thucydides, *The Peloponnesian War*, trans. by Rex Warner (Harmondsworth, U.K.: Penguin, 1954), 25. The recent literature is summarized and extended in *World Politics* 38 (October 1985), also published as Kenneth Oye, ed., *Cooperation under Anarchy* (Princeton: Princeton University Press, 1986). The framework used grows out of Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984); Robert Jervis, "Cooperation under the Security Dilemma," *World Politics* 30 (January 1978), 167-214; and Robert Keohane, *After Hegemony* (Princeton: Princeton University Press, 1984). Related arguments are made by Michael Taylor, *Anarchy and Cooperation* (New York: Wiley, 1976). Although the subject matter is the same as that treated in Hedley Bull, *The Anarchical Society* (New York: Columbia University Press, 1977), Bull's approach is different and his work is not cited in this literature. For arguments that Bull provides a better foundation for understanding international politics than does the work analyzed here, see Hayward Alker, Jr., "The Presumption of Anarchy in World Politics," and Richard Ashley, "Hedley Bull and the Anarchy Problematique," both in Alker and Ashley, eds., *After Realism: Anarchy, Power, and International Collaboration* (forthcoming).

[2] See Charles Lipson, "Bankers' Dilemmas: Private Cooperation in Rescheduling Sovereign Debts," *World Politics* 38 (October 1985), 200-225, for an analysis of non-state actors within this framework.

[3] O'Neill, "Game Theory and the Study of Deterrence of War," in Robert Axelrod, Robert Jervis, Roy Radner, and Paul Stern, eds., *Perspectives in Deterrence*, forthcoming.

and rigor that stem from formalization. But the formalization is limited: what is most important is the basic structure of the game and the analogies that are provided.[4] Game theory and Realism are generally compatible—both are structural, strategic, and rational—but each has its own vulnerabilities. Thus, some of my criticisms can be traced more to problems with Realism, others more to problems with game theory. But my focus is on the work that exemplifies the intersection of the two; it is not a full-blown critique of the entirety of either approach.

Drawing on the concepts of the iterated Prisoners' Dilemma (PD) and public goods,[5] the basic question posed by the recent work is how self-interested actors can cooperate in the face of anarchy and important conflicting interests. By taking two actors and positing that each of them has only two choices (cooperating with or defecting from the other), an interesting world emerges in which four outcomes are possible. They are preferred by the actor in the following order: first, the actor defects while the other cooperates (DC), thus allowing the former to gain an advantage; second, both actors cooperate (CC); third, both may defect (DD), thus producing competition; the fourth and worst outcome would be for the actor to cooperate while the other defects (CD) and thereby exploits him. What makes this configuration disturbing is that even if each side prefers CC to DD (and each knows that this is the other's preference), the result can be DD because each is driven by the hope of gaining its first choice—which would be to exploit the other (DC)—and its fear that, if it cooperates, the other will exploit it (CD).

When a good idea like this comes along, our exuberance at finding new insights leads us to extend and apply it widely, postponing a consideration of problems and limitations. To some extent this is functional: as Albert

[4] For contrasting evaluations of the potential of higher forms of game theory, see *ibid.*, and Duncan Snidal, "The Game *Theory* of International Politics," *World Politics* 38 (October 1985), 25-57. For an excellent general discussion, see Thomas Schelling, "What is Game Theory?" in Schelling, *Choice and Consequence* (Cambridge: Harvard University Press, 1984), 213-14.

[5] The relationship between PD and public goods is technical, complex, and subject to dispute. See John Conybeare, "Public Goods, Prisoners' Dilemma, and the International Political Economy," *International Studies Quarterly* 28 (March 1984), 5-22, and Russell Hardin, *Collective Action* (Baltimore: The Johns Hopkins University Press, 1982), 16-30. The essential similarity between the two that is relevant here is that the equilibrium solution is non-optimal. That is, in the absence of devices to avoid this outcome, individual self-interested rationality leads each actor to be worse off than he could have been if all players had acted differently. This is true even though public goods are characterized by non-rivalry of consumption and non-excludability—two dimensions that can be distinguished: see Duncan Snidal, "Public Goods, Property Rights, and Political Organizations," *International Studies Quarterly* 23 (December 1979), 532-66—in contrast to two-person PDs that have both rivalry and excludability. When large numbers are involved in a PD, as they are in the Tragedy of the Commons, excludability is precluded and the situation can be considered as one of a public good that has the characteristics of non-excludability but not non-rivalry. These distinctions are important in many analyses, but they are not central here.

Hirschman has noted, in many cases we would not start what will prove to be a fruitful enterprise if we were aware of the difficulties at the beginning.[6] But if we ignore the problems in the early stages, we may overreact and reject the entire approach when they become obvious later. It would be more productive if scholars simultaneously explored the potential of and the problems with the approach. Since the potential has been discussed elsewhere (see fn. 1), I will concentrate on the problems.

Three strengths that have made this research stimulating and important should be noted, however. First, it builds upon central characteristics of international politics—anarchy, the security dilemma, and the combination of common and conflicting interests. Second, the approach is parsimonious and lends itself to deductive theorizing. Third, it seeks to bring together the study of conflict and the study of cooperation, and tries to explain a wide range of phenomena encompassing both security and political economy. The deterring of exploitation may be as relevant to the stability of monetary systems as to arms control; strategies by which states can gain the benefits of mutual cooperation may be as important for nuclear postures as for international trade. Indeed, as the links to the theory of public goods indicate, the propositions produced should apply to a wide range of cases outside of international politics.

It is not a good sign, however, that prisoners confronted by a District Attorney do not behave as the model would lead us to expect.[7] In order to apply the framework, we have to *assume* many of the elements of world politics that in fact are most problematical. The actor's values, preferences, beliefs, and definition of self all are exogenous to the model and must be provided before analysis can begin.[8] If they were straightforward, the analysis of international politics would be much simpler. They therefore need to be investigated and explained, not taken as givens. Adoption of stereotypical Realist assumptions can supply some of the inputs the analysis requires, but I will argue that this is not completely satisfactory.

Choice and Context

Perhaps the most important limitation of the work on anarchy is that it looks at individual actors, their preferences, and their choices, and thus

[6] Hirschman's discussion of what he calls the "hiding hand" is in his *Development Projects Observed* (Washington, DC: Brookings Institution, 1967), 9-34.

[7] Brian Forst and Judith Lucianovic, "The Prisoner's Dilemma: Theory and Reality," *Journal of Criminal Justice* 5 (Spring 1977), 55-64.

[8] The model also assumes that states can fruitfully be considered as unitary actors. The debates over this issue, although important, are so well known that they will only be touched on here.

blinds us to the broader setting in which behavior occurs. Problems arise in a context and out of a history, and not all patterns are the product of careful or even conscious choice. Just as in most of our everyday lives we carry out routines, so states also continue established policies with little calculation or thought.[9] Each new day does not bring a new beginning; severe restrictions are placed on us by the expectations—including our own expectations about ourselves—that constitute the context within which we must behave.

When clear points of choice occur, they are often structured by the settings in which they arise. The timing of decisions and events is important and at least partially beyond a decision maker's control. For example, the production schedule for new nuclear submarines has periodically presented President Reagan with the choice of whether to abide by the unratified SALT II Treaty and destroy a usable but older missle-launching submarine, or to break at least part of this agreement. The date on which these decisions had to be made was in part an arbitrary product of industrial vagaries.[10] Furthermore, unexpected events may occur at the same time as the decision, and force themselves on the statesmen's feelings and calculations. Thus, by straining the NATO alliance, Reagan's air strike against Libya in April 1986 may have made it harder for him to break the treaty at the decision point six weeks later. In turn, this decision greatly increased the pressures on Reagan to take a tougher position at the next choice point in late 1986.[11] Similarly, the United States might have made more of an effort to cooperate with the Soviet Union in banning MIRVs had it not been simultaneously negotiating over ABMs. Another example is that American cooperation with the Mariel boatlift of refugees from Cuba was reduced by the accidents of a change in relevant laws—adopted for unrelated reasons—and a large simultaneous emigration from Haiti to the U.S.[12]

More generally, issues arise in particular historical contexts that shape

[9] James Rosenau stresses the role of habits in "Before Cooperation: Hegemons, Regimes, and Habit-Driven Actors in World Politics," *International Organization* 40 (Autumn 1986), 861-70.

[10] Michael Gordon, "Air Force's Delay Said to Keep U.S. to '79 Arms Limit," *New York Times*, August 29, 1986.

[11] Michael Gordon, "Reagan Declares U.S. Is Dismantling Two Nuclear Subs," *New York Times*, May 28, 1986; also see Gordon, "U.S. Still Divided on 1979 Arms Treaty," *New York Times*, April 23, 1986. When the U.S. did break the limits, some critics held that one motive was to show that the controversy over the Iran arms deal, which had just been revealed, would not inhibit the administration from acting decisively. Michael Gordon, "U.S. Exceeds Limit Set in 1979 Accord on Strategic Arms," *New York Times*, November 29, 1986.

[12] Jorge Dominguez, "Cooperating with the Enemy: U.S. Immigration Policies toward Cuba," in Christopher Mitchell, ed., *Immigration Policy and U.S. Foreign Relations with Latin America* (forthcoming). For additional examples, see Glenn Seaborg with Benjamin Loeb, *Stemming the Tide* (Lexington, MA: Lexington Books, 1987), 121, 158.

preferences and behavior. The operating incentives are given not only by the present circumstances, but also by how these circumstances came about. Where the players are is strongly influenced by where they have been.[13] Thus the U.S. decision to respond with force immedately to the seizure of the *Mayaguez* is to be explained largely by the fact that this event occurred shortly after the fall of Vietnam, when U.S. leaders believed that they had to act forcefully to show others that the previous defeat had not undermined American willingness to use force. Similarly, one reason why Britain could not cooperate with Germany and restrain Russia in 1914 was that she had done so in the previous crisis and feared that a repetition would destroy the Triple Entente. In other cases, defection or cooperation in one interaction will lead to more of the same. Cooperation can change the situation the actors face, and their beliefs, in ways that make further cooperation more likely. Perhaps more frequently, the process of mutual defection will lead each side to fear and dislike the other, and to develop distorted views of it, which cannot be quickly reversed if the other suddenly begins to cooperate.[14] Actors do not react merely to the *immediate* stimulus they face.[15]

We often talk of repeated plays of a Prisoners' Dilemma. But this formulation is misleading when the preferences and beliefs of the actors, and the nature of the game itself, change as it is played. What is at stake and the nature of the issue is defined over time, as actors develop their positions, in part in response to the positions taken by others. As one Senator explained his vote to permit the export of advanced arms to Saudi Arabia:

> The whole issue changed in the last week or 10 days. The media began to play it up as a question involving the president's ability to conduct foreign policy. . . . Had it not been for the media hype, the issue would have been strictly the arms sale. And I would have been very comfortable voting against it."[16]

[13] For a nice example taken from the pattern of social relations within a community of chimpanzees, see Jane Goodall, *The Chimpanzees of Gombe* (Cambridge: Belknap Press of Harvard University Press, 1986), 337-38.

[14] See the discussion of the "spiral model" in Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976), 62-84. For arguments that one or two instances of cooperation are not likely to be sufficient to break these spirals, see Charles Osgood, *An Alternative to War or Surrender* (Urbana: University of Illinois Press, 1962).

[15] For a discussion of a parallel in biological evolution, see Stephen Jay Gould, "Not Necessarily a Wing," *Natural History* 94 (November 1985), 12-25. Also see Gould, "Of Kiwi Eggs and the Liberty Bell," *Natural History* 95 (November 1986), 22-29. The argument that structures shape and limit later development is in sharp contrast to the standard evolutionary assertion that natural selection operates in a way that provides the best fit between organisms and their present environment.

[16] Quoted in Richard Fenno, Jr., "Observation, Context, and Sequence in the Study of Politics," *American Political Science Review* 80 (March 1986), 11-12.

The moves in a game can change it. Chicken can become Prisoners' Dilemma if each side's behavior leads the other to believe that being exploited would be worse than mutual defection. Schelling notes: "When a boy pulls a switch-blade knife on his teacher, the teacher is likely to feel, whatever the point at issue originally was, that the overriding policy question now is his behavior in the face of a switch-blade challenge."[17]

Other objections can be grouped under three headings, each of which has three components. The first set of questions concerns the actors' preferences, which need to be probed and explained. The second deals with the concepts employed by the framework, which may turn out to be ambiguous or troublesome. The third set involves the causes and consequences of the actors' perceptions, beliefs, and values. The representations of reality employed often beg crucial questions. Like many experiments, the models often gain internal validity at the cost of external validity: the order they impose is too rigid to catch the reality they seek to explain.

Preferences and Preference Orders

As noted, the central question for the work on anarchy is how cooperation is possible when actors are in a Prisoners' Dilemma—i.e., when they have the following preference order: exploiting or taking advantage of the other, mutual cooperation, mutual defection, and being exploited. The most important issue may be not what happens after these preferences have been established, but the preferences themselves. Much of the explanatory "action" takes place in the formation of the preferences; we cannot afford to leave this topic offstage. Three questions need to be addressed: How do we know what the actors' preferences are? What is the pattern of distribution of preferences over large number of cases? How are preferences established?

What Are the Actors' Preferences?

The first question is the most obvious one. How do we know whether a situation resembles a PD? The problem is even greater for those variants of the framework that call for cardinal rather than ordinal utilities.[18] The danger, of course, is that we will infer the actors' preferences from

[17] Thomas Schelling, *The Strategy of Conflict* (Cambridge: Harvard University Press, 1960), 265.

[18] This is true for Jervis (fn. 1).

their behavior. Actors rarely give complete statements of their preferences. In some cases we can use the method of "revealed preferences," but this technique can be used only when preferences are stable and consistent; if it is not to be tautologous, we must examine a large number of instances in order to test the preferences on behaviors that we did not use to derive the preferences in the first place.

Furthermore, while the standard PD model points to four possible outcomes that need to be ranked, decision makers may define the situation differently—most frequently by ignoring the possibility of mutual restraint. In 1914, for example, most leaders did not ask themselves whether they preferred peace to war because they did not think that peace could be maintained. In some cases, a lack of cooperation may be explained in significant measure by the actors' inattention to the possibility of such an outcome.

How Common Is Prisoners' Dilemma?

The second problem lies in determining the relative frequency of various games. How common is Prisoners' Dilemma? Harrison Wagner and George Downs and his colleagues note that states often fail to cooperate, not because they cannot surmount the PD, but because they are in Deadlock and prefer mutual defection to mutual cooperation.[19] For example, were the states in 1914 in PD or in Deadlock? (These are not, of course, the only alternatives.) Are we in PD with the Soviet Union today? Ronald Reagan is not alone in answering in the negative; much of the debate over whether American threats have deterred the Soviet Union or created unnecessary conflict is a disagreement about Soviet intentions and preference orders, not a dispute between two contending general theories.[20]

To put the questions in their most general form: How much of international conflict is caused by the states' inability to make and enforce binding agreements? How much of international politics is driven by this problem? The model of PD may be popular, not because it catches the most important dynamics of international politics, but because it is intriguing and lends itself to interesting manipulations. The "law of the instrument" may be at least partly at work. Furthermore, the model is congruent with the Anglo-American bias of seeing world politics as tragedy

[19] Harrison Wagner, "The Theory of Games and the Problem of International Cooperation," *American Political Science Review* 70 (June 1983), 330-46; George W. Downs, David M. Rocke, and Randolph M. Siverson, "Arms Races and Cooperation," *World Politics* 38 (October 1985), 118-46.

[20] Jervis (fn. 14), 84-113. Similarly, many of the arguments about whether the failure of detente was inevitable or a matter of errors can be phrased in terms of whether the relaxation of tensions was Pareto-superior to a high level of competition.

rather than as evil, of believing that most conflicts can be ameliorated for the good of all concerned.[21]

The empirical problems in answering these questions are relatively clear, although extremely difficult. Perhaps equally troublesome but less obvious are two conceptual difficulties. First, are we concerned only with situations in which the agreements cannot be enforced, or also with those in which no outside agency is available to force the actors to reach an agreement in the first place? Even where enforcement is available, as it is in domestic society, some mutually beneficial outcomes may be missed because of bargaining dynamics and miscalculations. Similarly, the dilemma of public goods is not that actors cannot ensure that others will live up to their commitments, but that it is not rational for anyone to make a commitment to contribute in the first place.

A second problem is that whether or not a situation is a Prisoners' Dilemma depends in part on how we define its boundaries. If we look at Japanese-American negotiations in the summer and fall of 1941, the framework does not apply: both sides preferred mutual defection (in this case, war) to the concessions that would have been necessary to reach agreement. Japan would rather fight than give up the effort to dominate China and South East Asia; the United States would rather fight than permit this to happen. But if we step back and ask why Japan sought its sphere of influence, anarchy and the Prisoners' Dilemma emerge as crucial. Domination was sought primarily as a means to the goal of autarky. Japanese leaders wanted the self-sufficiency that would go along with unhindered access to Asian markets and raw materials in order to become relatively immune from Western pressures.[22] Had it been possible for the West and Japan to make a binding agreement giving the latter economic freedom in Asia in return for renouncing the use of force, a bargain might have been struck.

Sources of Preferences

By taking preferences as given, we beg what may be the most important question on how they are formed. When and why do actors find exploitation particularly attractive, or the danger of being taken advantage of particularly unacceptable? When and why do decision makers see mutual cooperation as desirable? Economic theory treats tastes and prefer-

[21] See Arnold Wolfers, *Discord and Collaboration* (Baltimore: The Johns Hopkins University Press, 1962), chap. 15.

[22] James B. Crowley, *Japan's Quest for Autonomy: National Security and Foreign Policy, 1930-1938* (Princeton: Princeton University Press, 1966); Michael Barnhart, *Japan Prepares for Total War* (Ithaca, NY: Cornell University Press, 1987).

ences as exogenous. Analysis is therefore facilitated, but at the cost of drawing attention away from areas that may contain much of the explanatory "action" in which we are interested.[23]

Sometimes we can deduce preferences from the structure of the system, as Realism suggests. But even a structural theory of international politics as powerful as Waltz's has trouble producing many precise deductions.[24] As the endless arguments about the national interest remind us, only rarely can descriptions and prescriptions of what the state will or should prefer be drawn from its objective situation.[25] Conybeare is able to deduce a state's economic interests from trade theory;[26] but this inference only applies if we assume that the state is an actor. Such an assumption is often a valid guide when overriding issues of national security are at stake; but the relevant actors for economic issues are often classes, sectors, and groups, or even smaller and more numerous units. In the security area, the preference for protecting the status quo rather than retreating may be deducible from the external situation, but the choice for expansion is often the result either of preferences of important subnational groups or of internal bargaining.[27] More broadly, one does not have to be a Wilsonian or a Marxist to argue that the goals states seek, the costs they are willing to pay, and the instruments that are believed appropriate can be deeply affected by the nature of the state.

Transnational forces are a third general source of preferences. Thus, James Joll sees the prevalence of Social Darwinism in the early 20th century as a major cause of the preference for competition in world politics, and Van Evera argues that rampant nationalism strongly influenced the utilities that attached to various outcomes.[28] The maddeningly vague "spirit of the times" may help to account for the fact that, during the years before World War I, all the major powers opted for offensive military

[23] See the discussion of tastes in James March, "Bounded Rationality, Ambiguity, and the Engineering of Choice," *Bell Journal of Economics* 9 (Autumn 1978), 593-604.

[24] Kenneth Waltz, *Theory of International Politics* (Reading, MA: Addison Wesley, 1979).

[25] But see Stephen Krasner, *Defending the National Interest* (Princeton: Princeton University Press, 1978). Also see David Lake, "Beneath the Commerce of Nations: A Theory of International Economic Structure," *International Studies Quarterly* 28 (June 1984), 145-49, and Lake, "Power and the Third World: Toward a Realist Political Economy of North-South Relations," *International Studies Quarterly* 31 (June 1987), 221-28.

[26] John Conybeare, *Trade Wars* (New York: Columbia University Press, 1987).

[27] See Jack Snyder, *Myths of Empire*, forthcoming.

[28] James Joll, "1914: The Unspoken Assumptions," in H. W. Koch, ed., *The Origins of the First World War* (London: Macmillan, 1972), 307-28; Koch, "Social Darwinism as a Factor in the 'New Imperialism,'" *ibid.*, 329-54; Stephen Van Evera, "Why Cooperation Failed in 1914," *World Politics* 38 (October 1985), 80-117. For a related general argument on the source of preferences, see Aaron Wildavsky, "Choosing Preferences by Construction Institutions: A Cultural Theory of Preference Formation," *American Political Science Review* 81 (March 1987), 3-22.

strategies when defensive ones would have been more likely to make them reach their foreign policy goals.[29]

Preferences also stem from the ideologies and beliefs of individual decision makers. Some are "hard-line," others "soft-line" in dealing with an adversary. Whether these orientations hold across interactions with a range of adversaries we do not know, although some evidence indicates that they do.[30] The effect is that under circumstances in which some statesmen will believe that mutual cooperation is beneficial, others will see it as a trap. Thus, because of their beliefs about the nature of the adversary, Ronald Reagan and his supporters believe that many kinds of cooperation with the Soviet Union are likely to produce greater Soviet pressures on the West rather than further cooperation.

Herbert Simon's argument about Duverger's Law of the relations between voting rules and party systems applies to theorizing in international politics as well: "Most of the work is being done by propositions that characterize the utility function of the [actor] and his or her beliefs, expectations, and calculations."[31] Realism and game theory are of limited help here since the latter must assume the actors' preferences and utilities, and Realism's generalizations, although powerful, are often vague. Furthermore, these approaches imply that preferences are constant when, in fact, they change and thus pose a range of challenges for the anarchy framework.[32] First, and most obviously, a state's preferences change as one set of decision makers replaces another. Thus the dim prospects for arms control throughout much of the Reagan administration are explained by the President's belief that agreements that others saw as advantageous are in fact unacceptable. Game theory focuses on the prefer-

[29] Van Evera (fn. 28); Van Evera, "The Cult of the Offensive and the Origins of the First World War," *International Security* 9 (Summer 1984), 58-107; Jack Snyder, "Civil-Military Relations and the Cult of the Offensive, 1914 and 1984," *ibid.*, 108-46; Snyder, *The Ideology of the Offensive: Military Decision Making and the Disasters of 1914* (Ithaca, NY: Cornell University Press, 1984). For a rebuttal, see Scott Sagan, "1914 Revisited: Allies, Offense, and Instability," *International Security* 11 (Fall 1986), 151-76, and the exchange of letters between Snyder and Sagan, *ibid.*, 11 (Winter 1986/87), 187-98.

[30] For a discussion of this issue in the domestic context, see Robert Putnam, *The Beliefs of Politicians* (New Haven: Yale University Press, 1973).

[31] Simon, "Human Nature in Politics: The Dialogue of Psychology and Political Science," *American Political Science Review* 79 (June 1985), 298; also see William Riker, "The Heresthetics of Constitution-Making: The Presidency in 1787, with Comments on Determinism and Rational Choice," *American Political Science Review* 78 (March 1984), 1-16. In a different intellectual tradition, Adam Przeworski comes to a similar conclusion: "Marxism and Rational Choice," *Politics and Society* 14 (No. 1, 1985), 379-409.

[32] For a general discussion of changing preferences, see James March and Herbert Simon, *Organizations* (New York: Wiley, 1958), 141; Michael Cohen and Robert Axelrod, "Coping with Complexity: The Adaptive Value of Changing Utility," *American Economic Review* 74 (March 1984), 30-42; Barbara Farnham, "Value Conflicts and Political Decision-Making" (Ph.D. diss. in progress, Columbia University).

ences that are held at any given time, and therefore simply puts this question to one side; Realism implies that preferences come from the position of the state in the international system, and therefore implies that changes in leadership do not matter. Realism is less troubled by a second source of change: that produced by shifts in the external situation. On the simplest level, a previously desired outcome may become unacceptable because the state would no longer gain by it. Thus, for instance, states begin to resist mutual low tariffs as their competitive position deteriorates. Similarly, in the fall of 1961 American decision makers came to doubt the wisdom of their own nuclear test-ban proposal when they suspected "that the Russians might have drawn even with or even past us in some aspects of thermonuclear weapons."[33]

In other cases, preferences change through the process of the interaction itself. When two states are hostile, one will often assume that anything the other urges must be bad; an outcome that was previously favored will be viewed with skepticism if the adversary endorses it. Similarly, the fact that a state takes a strong interest in an issue can lead the adversary to develop a contrary preference. The conflict process then *generates* interests and preferences rather than being produced by them.[34] It would otherwise be impossible to explain why the United States supports Jonas Savimbi in Angola, or refuses to admit refugees from Cuba. Unlike other communist regimes, Cuba *wants* its internal opponents to leave; so the U.S., counter to its normal policy of accepting such people, keeps them out.[35] Although this interdependence between states' preferences is compatible with Realism, it leads to interests and maneuvers that complicate, if not contradict, the normal prescriptions produced by the anarchy framework. For example, splitting up a large transaction into a series of small ones may be ineffective in these circumstances.

Experience and knowledge can also change preferences. Although such processes account for many important outcomes, they are exogenous to game theory and the anarchy framework. As actors participate in an arrangement, they may come to see that the consequences are quite different from those they expected. Thus the experience of the follow-up conferences on the Helsinki Agreement seems to have convinced the Soviets that the discussion of human rights, while annoying, does not

[33] Glenn Seaborg, *Kennedy, Khrushchev, and the Test Ban* (Berkeley: University of California Press, 1981), 120.

[34] Of course, each state wants to limit the other's power and influence and will therefore try to see that any actor the adversary is supporting does not win; but from this we cannot deduce preferences for specified countries or factions to prevail.

[35] Linda Greenhouse, "U.S. Assailed Again on Curbing Cuban Immigrants," *New York Times*, September 27, 1986.

do as much harm as they once believed.[36] In other cases, the emergence of new scientific information may be the source of change. The U.S. position in the test-ban negotiations was altered when people realized that testing might take place inside a large cavern, thus muffling the blast and undercutting the verification techniques in which the West previously had faith. New information or new beliefs can also lead to preferences for cooperation. Robert Rothstein argues, for example, that North-South conflict over commodity pricing decreased as the developing states came to see that their previous proposals for price stabilization might not increase their income, as they had initially believed.[37] Negotiations proceed not only by bargaining, but also by persuasion; in some cases, that leads to solutions that neither side had previously thought of and that each comes to prefer to its original proposal.[38]

Finally, preferences may be unstable. Because the intellectual problems are great, key decisions are often difficult, and continued thought may produce shifting evaluations. Although such processes are difficult to generalize about, most people know them from their own experience: deciding what car to buy, what job to take, or whom to marry. Most sharply, decision makers may come to prefer mutual cooperation to mutual defection, or vice versa. Thus, the Soviet-American negotiations for arms control in the Indian Ocean broke down largely because the Carter administration, after first favoring an agreement, came to believe that it would not be in the American interest. Similarly, in the early 1960s, the United States favored land-mobile missiles, then (in SALT I) proposed prohibiting them, then favored them as stabilizing, and has recently again called for banning them. The same kind of changes have occurred in the U.S. stance toward antisatellite weapons. Those shifts can only in part be explained by changes in technological possibilities or Soviet activities.[39]

[36] John Maresia, "Helsinki," in Alexander George, Philip Farley, and Alexander Dallin, eds., *U.S.-Soviet Security Cooperation: Achievements, Failures, Lessons* (New York: Oxford University Press, forthcoming 1988). Also see Raymond Garthoff, *Detente and Confrontation* (Washington, DC: Brookings Institution, 1985), 480. For an argument that people's preferences are often formed by their behavior, see Daryl Bem, "Self-Perception Theory," in Leonard Berkowitz, ed., *Advances in Experimental Social Psychology*, VI (New York: Academic Press, 1972); for an application to international politics, see Deborah Larson, *The Origins of Containment* (Princeton: Princeton University Press, 1985).

[37] Rothstein, "Consensual Knowledge and International Collaboration: Some Lessons from the Commodity Negotiations," *International Organization* 38 (Autumn 1984), 732-62.

[38] See Mary Parker Follett, *Dynamic Administration: The Collected Papers of Mary Parker Follett*, H. C. Metcalf and L. Urwick, eds. (New York: Harper & Row, 1942); Richard Walton and Richard McKersie, *A Behavioral Theory of Labor Negotiations* (New York: McGraw-Hill, 1965), 126-83; Ernst Haas, *Beyond the Nation-State* (Stanford, CA: Stanford University Press, 1964), 86-125.

[39] For a discussion of the ways in which changes in technology influenced American attitudes toward antisatellite weapons, particularly by providing satellites with offensive as well

In summary, the research on cooperation under anarchy assumes the actors are in PD, which may not be correct. It further takes the actors' preferences as given and ignores how, why, and when they change. Game theory cannot help here; Realism's analyses of these questions are limited at best.

Concepts

The concepts employed by the anarchy framework seem unproblematic at first glance. Cooperation and defection, offense and defense, and power are fairly standard concepts, but the requirements for rigorous analysis within a game-theoretic approach demand that we squeeze some of the richness out of these terms. The admirable gain in precision may have to be purchased at an unacceptably high price.

Cooperation and Defection

The concepts of cooperation and defection are crucial to the framework. These terms work well for a laboratory Prisoners' Dilemma, but most situations are more complex. To start with, are these the only two alternatives? Perhaps we should think not of a dichotomy, but of a continuum. But can we add this measure of realism without sacrificing the parsimony and deductive power of the theory? Furthermore, can most alternatives really be arrayed along such a continuum? Some policies express a high degree of both cooperation and defection simultaneously; others—in some instances, the policy of isolation—express neither.[40] In many cases the most interesting choices are not even on this continuum.

Concrete questions often arise about how to characterize any individual outcome or policy. Sometimes these questions come down to whether the glass is half full or half empty, but even then the answer is not inconsequential. An example of an outcome is the "chicken war" between the United States and the E.E.C. in the early 1960s, which Conybeare sees as mutual defection because Europe raised the tariff on frozen poultry and the U.S. retaliated by raising tariffs on some European products.[41] But one can argue that what is more significant than the deviation from pure cooperation is the fact that the dispute was held to one round of action

as defensive military capabilities, see Paul Stares, *The Militarization of Space* (Ithaca, NY: Cornell University Press, 1985), and Steve Weber and Sydney Drell, "Cooperation and Discord in the Militarization of Space: U.S. Strategy, 1960-1985," in George, Farley, and Dallin (fn. 36).

[40] Thus it is not surprising that quantitative studies find that conflict and cooperation are not always inversely related to each other.

[41] Conybeare (fn. 26).

and counteraction. It did not produce a spiral of conflict, and relations between the countries were not embittered. An example of an ambiguous military policy is the question whether the U.S. cooperated or defected in its strategic arms procurements between the mid-1960s to the late 1970s. On the one hand, the United States, unlike the Soviet Union, did not deploy additional missiles. On the other, it did make major qualitative improvements, most notably in the form of multiple warheads.

How we judge a policy may also depend on the time span we examine. Defection in one instance can produce mutual cooperation over the longer run. Indeed, one implication of the theorizing about anarchy is that cooperation is enforced by the possibility of defection. Thus, President Reagan argues that his policy will induce Soviet cooperation, which will then be mutual. If he turns out to be correct, is it more useful to call the whole policy one of cooperation, or to divide it into discrete rounds and label the initial American behavior as defection? And if the Soviets have been conciliatory in order to weaken the West, and will ultimately return to an uncooperative stance from an improved position, should the policy not be characterized as defection? A simpler version of this ambiguity is revealed in the Cuban missile crisis. This episode is often considered an American victory; in fact, *mutual* cooperation greatly increased in the months that followed. If we take an even longer perspective, however, we might label the outcome as mutual defection, because of the Soviet arms build-up that was caused partly by the humiliation in Cuba. In game theory or in the laboratory, we can define what each play of the game is. But in the situations we are trying to analyze, these distinctions, and the answer to the question whether the behavior is seen as cooperative or not, are subjective if not artificial.

In other cases, the very meaning of cooperation is unclear. It usually denotes doing what the other actor prefers, but how do we characterize a response that is undesired but is designed to benefit the other, and/or has that effect? How do we classify behavior when one side desires a high degree of friction with the other and the other responds with the sought-for hostility? What do we say about cases in which neither side thinks about the impact on the other? Does the notion of defection imply something more than, or different from, noncooperation? The labels may squeeze out much of the reality that we are trying to catch. A recent example displays several of these ambiguities. In April 1986, the United States asked East Germany to restrict the travel of Libyan diplomats who might be aiding terrorism in the West. In response—or what they asserted to be response—the East Germans demanded that all diplomats except those of Britain, France, and the U.S. show their passports as they

crossed from East to West Berlin. (Previously, only special passes issued by the East German Foreign Ministry had been required.) The U.S. took this action, not as cooperation, but as "a clearly unfriendly negative act that is intended to undermine the four-power status" of Berlin.[42] This interpretation is not unreasonable, but neither is it compelled by the facts.

The problem of the subjective and political nature of definitions of cooperation is compounded in the cold war because the two sides have very different perspectives. Most Americans think about cooperating with the U.S.S.R. to maintain the status quo or to bring about mutually beneficial changes, particularly in arms control. But the Soviets not only consider the status quo in the sense of the current distribution of influence as disadvantageous, but believe that it cannot be maintained because it is being moved by the forces of history. Thus, for the Soviets, CC consists of managing the changes in a peaceful way. While the United States would steadily lose ground, the Soviets argue that Washington should accept this process because the alternative would be even worse.[43]

Even when such problems are absent, it is easy to overlook the fact that what is deemed a defection is in part rule-governed. The United States—and most of the world—considered the Soviet emplacement of missiles in Cuba in 1962 as strongly noncooperative; the Soviet arms build-up of the late 1960s, which had a much greater influence on the military balance, was seen as less of a defection. Although what Waltz calls "internal balancing" may not be welcomed by others, it is accepted as a normal part of international politics. It is seen as legitimate; indeed, coercive attempts to block it would be perceived as illegitimate and highly threatening. These distinctions cannot be understood by measuring the objective harm done to the state. The fact that the actors usually give little thought to the rules does not reduce their importance in defining the crucial concepts we use.

Some of these difficulties are summarized by the question a student raised when I played a version of multiperson Prisoners' Dilemma in class: what, she asked, are we to cooperate *about*? For decision makers, the question is never cooperation or defection, but rather what goals to seek and the tactics that will be most apt to reach them. Of course it is not illegitimate to impose our categories on the behavior of actors—and the notions of cooperation and defection are hardly foreign to them—but

[42] James Markham, "Allied Diplomats Defy East German Controls," *New York Times*, May 28, 1986. For parallel discussion of an important case in the late 18th century, see Paul Schroeder, "Old Wine in New Bottles: Recent Contributions to British Foreign Policy and European International Politics, 1789-1848," *Journal of British Studies* 26 (January 1987), 10.

[43] See Garthoff (fn. 36), 38-50, 1069. The Japanese had a similar conception of cooperation with the British in China in the 1930s: see Paul Haggie, *Britannia at Bay* (Oxford: Clarendon Press, 1981), 126.

when actors define their problems and choices quite differently from the theoretical model, we may be forcing disparate behavior into an inappropriate framework.

Offense and Defense

Questions can also be raised about the concepts of offense and defense. Both logic and evidence indicate that the prospects for cooperation are increased to the extent that defensive military systems and strategies can be distinguished from, and are stronger than, offensive ones.[44] In principle, the difference between offense and defense is clear: the former involves the ability to attack the other and seize his territory; the latter the ability to repel such an attack and protect one's own territory. The issue of whether attacking or defending is preferable comes down to whether a decision maker who believed that war was inevitable would prefer to attack the other side, as in 1914, or be on the defensive while the other side attacked—which, as the history of World War I shows, would have been a better strategy.[45]

But questions arise. Even if it is possible to say whether the offense or the defense has the advantage, can one distinguish between offensive and defensive weapons and strategies? Much of the interwar disarmament efforts were devoted to doing so. The fact that statesmen worked hard on this question indicates that they believed that a positive answer was attainable; that such efforts yield success only occasionally indicates that the endeavor may be doomed—although in the interwar case the reason may have been the refusal to forgo offensive options rather than the inability to make the relevant distinctions. Similarly, George Downs and his colleagues find that arms races are not terminated by the shift to defensive systems[46]—which indicates that statesmen are either unwilling or unable to avoid threatening others while seeking to maximize their own security.

Can the concepts of offense and defense be applied to nuclear weapons? A frequent argument is that the common-sense definitions have to be turned on their heads. That is, offense is the ability to take one's cities out of hostage; conversely, the ability to destroy the other side's population and other values, previously associated with the offense, is now con-

[44] See Jervis (fn. 1), at 186-214; George Quester, *Offense and Defense in the International System* (New York: Wiley, 1977); Jack Levy, "The Offensive/Defensive Balance of Military Technology: A Theoretical and Historical Analysis," *International Studies Quarterly* 28 (June 1984), 219-38; Van Evera (fns. 28 and 29); Snyder (fn. 29); Barry Posen, *The Sources of Military Doctrine* (Ithaca, NY: Cornell University Press, 1984); Weber and Drell (fn. 39).

[45] See the literature cited in fn. 29 above.

[46] Downs, Rocke, and Siverson (fn. 19).

sidered defensive (as long as the other side has a similar capability) because such an act could be credibly threatened only as retaliation for the other's attack. The obvious implications are that superpower strategies designed to deter war by developing counterforce capabilities exacerbate the security dilemma and make cooperation more difficult, and that countervalue targeting would permit mutual security. But such paradoxical reasoning has not been accepted by all analysts, with the result that there is no consensus about how the concepts fit the current situation.

Even if the distinction between offense and defense works, at least to some extent, in the military arena, can it be transferred to other dimensions of international relationships?[47] To find the answer, one would have to ask under which conditions bandwagoning as opposed to balancing is likely to take place.[48] The former dynamics make the offensive potent; the latter support the defensive. When initial gains are expected to create positive feedback, each state will have strong incentives to defect, irrespective of whether its ultimate intentions are aggressive; moving first may endanger the other, but it is necessary in order to protect the state. As a result, mutual restraint will be difficult to maintain even if both sides are satisfied with the status quo. By contrast, when balancing prevails and gains by one side call up counteracting pressures, states can afford to await developments, making major efforts only on the relatively rare occasions when another power poses a direct threat. Cooperation should thus be more prevalent when bandwagoning is neither feared nor hoped for.

The most obvious example of bandwagoning—or at least the expectation of it—is the "domino theory." When a defeat in one country is expected to have major and deleterious consequences for the state's influence in other areas, then even minor threats must be met swiftly and firmly. On the other hand, when one side believes that the other side's local victory is likely to be contained—for example, by the efforts of neighbors who are alarmed by the new threat—then it can afford to try cooperating with the other because the costs of being taken advantage of are relatively low. The political defensive is potent; the security dilemma

[47] For one effort at doing so, see Robert Jervis, "From Balance to Concert: A Study of International Security Cooperation," *World Politics* 38 (October 1985), 58-79, at 62-64. For a further discussion, see Jervis, "Cooperation Under Anarchy: Problems and Limitations," in Alker and Ashley (fn. 1).

[48] Arnold Wolfers, "The Balance of Power in Theory and Practice," in Wolfers (fn. 21), 122-24; Waltz (fn. 24), 125-28; Stephen M. Walt, "Alliance Formation and the Balance of World Power," *International Security* 9 (Spring 1985), 3-43, and *The Origins of Alliances* (Ithaca, NY: Cornell University Press, 1987); Jack Snyder and Robert Jervis, eds., *Strategic Beliefs and Superpower Competition in the Asian Rimland*, forthcoming.

is ameliorated; and a state can gain a high degree of protection without proportionately undermining the interests of its adversary.

Arguments like these need much more development before we can be sure that the distinction between offense and defense is possible and helpful in the political arena. What we have so far are essentially exploratory probes; the results, while suggestive, are hardly definitive. Furthermore, we have not even tried to apply such an extension to the realm of international economic relations.

The Nature of Power

A third conceptual difficulty is created by the fact that in many encounters the main stake is power—a notoriously difficult concept. But in almost all formulations, power in international politics is relative rather than absolute.[49] When states think about the possibility of an armed conflict, they have to judge how their forces compare to those of their adversaries; the absolute sizes of the forces on each side are irrelevant. This is true for many political conflicts as well. Knowing how much leverage one state has over another tells statesmen and analysts very little unless they also know how much leverage the other state has. Thus, it can be rational for statesmen to act in ways that reduce the absolute level of benefits they receive.

This view is not only compatible with Realism, but is embedded in it. But it shows that, as long as power is central, an element of inherent conflict will be involved, thus complicating statesmen's attempts to establish cooperation and undercutting some of the prescriptions that can be deduced from the anarchy framework. When states care primarily about maximizing their power advantage and not about absolute gains and losses, many of the strategies and conditions that should lead to cooperation in a Prisoners' Dilemma no longer produce this result.[50] Thus Robert

[49] Nuclear weapons may have changed this, as is indicated by the title and essays in Bernard Brodie et al., *The Absolute Weapon* (New York: Harcourt Brace, 1946).

[50] The implications of the fact that states often seek relative rather than absolute gains have been discussed by Arthur Stein, "The Hegemon's Dilemma," *International Organization* 38 (Spring 1984), 355-86. Also see Joanne Gowa, "Anarchy, Egoism, and Third Images: *The Evolution of Cooperation* and International Relations," *International Organization* 40 (Winter 1986), 176-77, and Joseph Grieco, "Distributional Uncertainty and the Realist Problem of International Cooperation," paper presented to the 1986 APSA meeting. For a discussion of the issues and a summary of the experimental literature, see Deborah Larson, "Game Theory and the Psychology of Reciprocity" (unpub., Columbia University), 25-31. More generally, Fred Hirsch, in *Social Limits to Growth* (Cambridge: Harvard University Press, 1976), brilliantly demonstrates that many goods in society are "positional"—that is, they are inherently competitive. Only a few people can be at the top of an established hierarchy of power or prestige. Certain goods—such as living in an uncrowded area—cannot be shared with large numbers of others. More fundamentally, in many aspects of life we judge how well we are doing by comparing ourselves to others.

Axelrod points to the irony that the strategy of reciprocity (Tit-for-Tat) that proves so effective in computer tournaments of PD can never win any individual game.[51] The strategy works well, however, because it accumulates large numbers of points when matched against fellow cooperators and is not terribly badly exploited when matched against more competitive strategies. But if the main rewards in international politics are for relative gains, a strategy that can, at best, only tie will not be attractive.[52]

The conditions under which states seek to maximize their relative as opposed to their absolute gains need more exploration.[53] These concerns are likely to be greater in the security area than in international economics, but they are present in the latter as well, especially because military and economic strengths are closely linked. Third-world states often seek greater economic equality with the developed countries as well as absolute economic growth. Furthermore, the drive to increase their political power vis-à-vis the developed states would make relative gains even more important.[54]

Even among allies, concern for relative gains is rarely completely absent. For example, would the United States gain or lose (or both) if Japan made new breakthroughs in microelectronics? The U.S. would benefit by importing better goods, but would fall further behind Japan in various technologies. This, in turn, might enable Japan to keep widening the productivity gap with America. More than national pride is involved, especially since no one can guarantee that Japan will maintain its political orientation in the future. Indeed, Stein has shown that the "hegemon's dilemma" operates in the economic arena because a dominant state that cares only about absolute gains is more likely to be overtaken by other powers.[55] Aaron Friedberg has similarly demonstrated the importance of this issue in the debate about the decline of Britain's economic power at the turn of the century. Those who were least concerned stressed the continuing growth of the British economy; those who felt the situation was dangerous concentrated on the fact that other economies were growing even faster.[56]

[51] Axelrod (fn. 1).

[52] See the discussion of strategies in Axelrod's computer tournament in Roy Behr, "Nice Guys Finish Last—Sometimes," *Journal of Conflict Resolution* 25 (June 1981), 289-300.

[53] The relevant literature from experimental psychology is summarized in Larson (fn. 50), 28-29.

[54] See Robert W. Tucker, *The Inequality of Nations* (New York: Basic Books, 1977); Stephen Krasner, *Structural Conflict* (Berkeley: University of California Press, 1985).

[55] Stein (fn. 50).

[56] Friedberg, *The Weary Titan: Britain and the Experience of Relative Decline, 1895-1905* (Princeton: Princeton University Press, forthcoming 1988), chap. 2.

Waltz's argument raises similar difficulties. He denies that states seek to maximize either absolute or relative power. Instead, they try to maximize their security.[57] But since the state's security can be influenced—positively or negatively, depending on the situation—by the power and security of others, this concept is likely to be inherently interactive. Unlike the desire to maximize relative gains, this may increase cooperation, but the result still will complicate the standard proto-game theory formulations.

Perceptions, Values, and Self-Interest

The final set of problems concerns psychology, beliefs, and values. In increasing order of conceptual difficulty, the issues are the psychological impediments to cooperation, the role of values and the autonomy of beliefs, and the question of whether narrow self-interest can explain most international behavior. Realism has generally ignored the decision-making level of analysis; game theory can incorporate the empirical findings in this area only at the cost of parsimony and a revision of many current conclusions. To understand when cooperation occurs, we need to see how different beliefs and values can affect the actors' evaluation of the outcomes. A related question is whether self-interest is defined coolly and atomistically or is driven by passions and care for what happens to others. If the latter is the case, the important factors will be outside of, if not denied by, the anarchy framework.

Psychological Impediments to Cooperation

Robert Keohane has noted that cooperation may be explained in part by the fact that people can operate with only bounded rationality; statesmen need to conserve cognitive resources, and shared norms and principles can be extremely useful in easing the burdens of prediction, choice, and coordination.[58] But cooperation is decreased by other aspects of the way people think. People do not mechanically reciprocate conciliation and defection. Instead, their behavior is mediated by their analysis of what the other did and why it did it. The interpretation of others' action is rarely self-evident, but it is almost always important. Most behavior is ambiguous; even more so are the underlying intentions. Since actors' responses are linked to their predictions of how the other will react to alternative policies they can pursue, their attributions of the causes of the

[57] Waltz (fn. 24); see also Waltz, "Reflections on *Theory of International Politics*: A Response to My Critics," in Robert O. Keohane, ed., *Neorealism and Its Critics* (New York: Columbia University Press, 1986), 334.

[58] Keohane, *After Hegemony* (Princeton: Princeton University Press, 1984), 110-32.

other's behavior will be crucial. Thus, how they react is influenced by their inferences as to whether the other intended the results that were produced and whether the behavior is best explained by the transient situation the other was in or by its enduring dispositions.[59] Furthermore, since international politics is an interactive process, a statesman's understanding of the other's behavior is influenced by how he thinks his own state is behaving toward the other. Indeed, perhaps the most important psychological factor that interferes with cooperation is that statesmen—and people in their everyday lives—greatly underestimate the extent to which their actions threaten or harm others. They think they are cooperating when an objective observer would say that they are, at least to some extent, defecting.

Part of the reason is that most statesmen are only dimly aware of the security dilemma. When they are peaceful, they think that their intentions are clear and that others will not be threatened by the measures they are taking in their own self-defense. Former Assistant Secretary of Defense Richard Perle once said that if we are in doubt about Soviet intentions, we should buy arms: if the Soviets are aggressive, the build-up will be needed, and if they are not, the only consequence will be wasted money. Similarly, when U.N. troops were moving toward the Yalu, Secretary of State Acheson stated that there was no danger that the Chinese would intervene out of self-defense because they understood that the U.S. was not a threat to them.[60]

Statesmen who think well of themselves generally believe that their actions are compatible with the reasonable interests of others. For example, Raymond Garthoff shows that, during the period of detente, Americans—leaders and general public alike—believed that their country was restrained and cooperative toward the U.S.S.R.; actually, the U.S. continued to seek unilateral advantage by freezing its adversary out of the Middle East, courting China, and developing advanced military technology.[61] As a result of this faulty self-image, any hostile reaction by the adversary is likely to seem unprovoked.

This problem is compounded by a second and better-known bias: states tend to overestimate the hostility of others and will often see as de-

[59] For further discussion of the inference processes and evidence of their importance drawn from experiments and case studies, see Jervis (fn. 14), 32-48, and Larson (fn. 50).

[60] Daniel Yergin, " 'Scoop' Jackson Goes for Broke," *Atlantic Monthly* 233 (June 1974), 81. (The same error is made, even more crudely, by Secretary of Defense Weinberger in *Annual Report to the Congress, FY 1988* [Washington, DC: Government Printing Office, 1987], 16.) Acheson's views are presented in John Spanier, *The Truman-MacArthur Controversy and the Korean War* (New York: Norton, 1965), 97. Similar examples are discussed in Jervis (fn. 14), 67-76, and Seaborg (fn. 12), 30-31.

[61] Garthoff (fn. 36). Also see Van Evera's discussion (fn. 28) of the role of nationalism in preventing statesmen from objectively gauging the behavior of their own states.

fection actions that a disinterested observer would record as at least partly cooperative. Dulles's view of Khrushchev's arms cuts in the mid-1950s is one such example; Reagan's view of most Soviet arms proposals may be another.[62]

These two biases often operate simultaneously, with the result that each side is likely to believe that it is cooperating and that the other has responded by defecting. Thus, the United States says that while it has been restrained, the Soviet Union has accumulated many more arms than it needs for deterrence. The Soviets probably have a similar perception of the American defense posture and may think they have limited their own arms. The breakdown of detente reveals the same pattern of perceptions (although that may not provide the primary explanation for the inability of the superpowers to maintain cooperation). To take just one example, when Brezhnev told Nixon in the spring of 1973 that the status quo in the Middle East was unacceptable, and when Gromyko later warned that "the fire of war [in the Middle East] could break out onto the surface at any time," they may well have thought that they were fulfilling their obligations under the Basic Principles Agreement to consult in the event of a threat to peace. The Americans perceived the Soviets to be making threats and, by failing to restrain the Arabs or to notify the U.S. of the impending attack, to be violating the spirit of detente.[63] The same effect often appears during bargaining when each side thinks that the concessions it has made are greater than those it has received.[64]

Furthermore, when statesmen realize that the other side has cooperated, they often believe that it did not have much choice. Thus, decision makers underestimate the ability of others to defect, and therefore frequently believe, incorrectly, that they can get away with some exploitation. This factor played a role in the deterioration of the recent detente as well as in the cooling of relations between England and France in the 1840s.[65] The Soviet Union in the 1970s, and both sides in the earlier in-

[62] See the classic essay by Ole Holsti, "Cognitive Dynamics and Images of the Enemy: Dulles and Russia," in David Finlay, Ole Holsti, and Richard Fagen, *Enemies in Politics* (Chicago: Rand McNally, 1967), 25-96. Michael Sullivan, *International Relations: Theories and Evidence* (Englewood Cliffs, NJ: Prentice-Hall, 1976), 45-46, questions the links between Dulles's beliefs and American behavior, however.

[63] Gromyko is quoted in Galia Golan, *Yom Kippur and After* (London: Cambridge University Press, 1977), 68. The treatment of the 1973 war is a good litmus test for one's views on detente: compare, for example, the discussions in Harry Gelman, *The Brezhnev Politburo and the Decline of Detente* (Ithaca, NY: Cornell University Press, 1984), chap. 4; Garthoff (fn. 36), chap. 11; and Alexander George, *Managing U.S.-Soviet Rivalry* (Boulder, CO: Westview Press, 1983), chap. 7.

[64] For experimental support for this proposition, see Michael Ross and Fiore Sicoly, "Egocentric Biases in Availability and Attribution," *Journal of Personality and Social Psychology* 37 (March 1979), 322-36.

[65] Roger Bullen, *Palmerston, Guizot, and the Collapse of the Entente Cordial* (London: Athlone, 1974), 81, 93.

stance, thought they could safely act against the other side's interests because they exaggerated the constraints inhibiting the rival's retaliation.

Because of biases like these, analyses that assume both sides to be perceiving each other accurately are likely to be incorrect; strategies that are based on this assumption are likely to be ineffective. Axelrod shows that the strategy of Tit-for-Tat works quite well when there is a one-percent error rate in the correct identification of the other's behavior.[66] But such a figure is drastically lower than that which can be expected in political interactions; George Downs and his colleagues have shown that when the error rate is higher, this strategy is not likely to yield stable cooperation.[67] It is partly for this reason that Larson argues that Charles Osgood's GRIT (Graduated Reciprocation in Tension-reduction), which makes a major effort to break through the adversary's perceptual biases, may induce more cooperation than Tit-for-Tat.[68] Strategies that are not tightly conditional on what the other does, and that do not require immediate and matching reciprocation, may produce a change in the other's attitudes, and so may be more effective than game-theoretic analyses would suggest. Strict reciprocity may fare less well than expected. Put in slightly different terms, cognitive biases decrease transparency, thus making regimes harder to establish and maintain.

Cooperation is more likely to come about when actors correct for these conflict-inducing biases or are willing to tolerate a higher level of perceived cheating by the other side. But such conditions and strategies also increase the chances that the other side will, in fact, cheat. One of the reasons why Israel was taken by surprise in 1973 was that its decision makers were highly sensitive to the security dilemma and believed that the most probable cause of war would be Egyptian preemption growing out of an incorrect belief that Israel was about to strike. They therefore interpreted Egypt's behavior preceding the war as evidence, not that Egypt was defecting, but that its decision makers thought Israel might be planning to do so. As a result, Israel did not respond by increasing its own military readiness.[69] The problems of misperception thus heighten the statesman's dilemma in deciding whether to cooperate or not.

[66] Axelrod (fn. 1), 182-83.

[67] Downs, Rocke, and Siverson (fn. 19), 133-34. Also see Jonathan Bendor, "In Good Times and Bad: Reciprocity in an Uncertain World," *American Journal of Political Science* 31 (August 1987), 531-58. Downs and Rocke discuss the implications of this finding in "Tacit Bargaining and Arms Control," *World Politics* 39 (April 1987), 297-325.

[68] Deborah Larson, "Crisis Prevention and the Austrian State Treaty," *International Organization* 41 (Winter 1987), 30-34.

[69] Janice Stein, "Calculation, Miscalculation, and Conventional Deterrence II: The View from Jerusalem," in Robert Jervis, Richard Ned Lebow, and Janice Stein, *Psychology and Deterrence* (Baltimore: The Johns Hopkins University Press, 1985), 60-88. This is part of the broader difficulty statesmen face in deciding whether the other side is an aggressor who must

Strategies that are robust in the face of misperception are extremely valuable, but it is far from certain that they exist.[70] Stable cooperation is most likely to result when the decision makers' preconceptions provide an accurate fit with what the other side is like; that may be as much the product of luck, however, as of sensitivity and statesmanship. It is clearly important to determine the extent to which strategies that would work well when information is accurate can also serve in a world permeated by ambiguity and strong perceptual and decision-making biases. Neither scholars' analyses nor statesmen's policies can safely be based on the assumption that either side understands the other.

Beliefs and Values

Expected utilities are the valuation an actor places on a course of action or outcome; they involve both estimates of consequences and judgments about intrinsic worth. Contrary to the implications of many Realist writings, these estimates and judgments are not objective, and they should not be accepted without investigating their formation, as is done in most of the work on game theory and anarchy. In many cases, what is crucial in determining whether the actors cooperate is their beliefs about the effectiveness of alternative policies—beliefs that often prove to be either lacking in evidentiary support before the fact and/or wrong after the fact. The most obvious and consequential example is the great exaggeration by decision makers of the efficacy of the offensive before World War I, which meant that a crisis was likely to lead to war.

The general question raised by this case is the extent to which (and the circumstances under which) the main impediments to cooperation are rooted in potentially malleable beliefs about the situation rather than in its structure.[71] For example, many historians have argued that there could have been much greater cooperation between Britain and Russia in the 19th century if the former had not greatly exaggerated the hostility and capability of the latter.[72] More generally, decision makers are often un-

be met with firmness (if not force), or a more reasonable state that can be conciliated or, to use the older term, appeased. See Jervis (fn. 14), chap. 3.

[70] *Ibid.*, 109-13; Richard Ned Lebow, "The Deterrence Deadlock: Is there a Way Out?" in Jervis, Lebow, and Stein (fn. 69), 180-202.

[71] See Ernst Haas, "Why Collaborate? Issue-Linkage and International Regimes," *World Politics* 32 (April 1980), 357-405; Robert Rothstein (fn. 37); Robert Jervis, "Security Regimes," *International Organization* 36 (Spring 1982), 359-60, 373-75; Jack Snyder, "Perceptions of the Security Dilemma in 1941," in Jervis, Lebow, and Stein (fn. 69).

[72] See, for example, Norman Rich, *Why the Crimean War?* (Hanover, NH: University Press of New England, 1985), and R. W. Seton-Watson, *Disraeli, Gladstone, and the Eastern Question* (New York: Norton, 1972).

aware of the security dilemma, and therefore act in ways that compound it. Is it possible to alter such perceptions, and thus alter behavior? One reason why the Kennedy and Carter administrations were favorably predisposed toward arms control was that many of the top officials believed that arms races were often driven by action-reaction cycles. Similarly, can perceptions of the relative efficacy of offense and defense be altered without changing technology? If military leaders are driven to favor the offensive because of organizational interests, and if civilian leaders are also disposed to underestimate the power of the defensive,[73] then perhaps the security dilemma and the resulting DD is more a creature of biases and domestic interests than of the structure of the international system.

This question arises with the current concern that World War III is most likely to start through "crisis instability" approximating that which prevailed in 1914. The vulnerability of command, control, and communications facilities might lead statesmen to believe that, even though any nuclear war would be dreadful, there would be real advantages to striking first if the choice were between doing so and being attacked. On closer examination, however, it does not appear that even a well-executed first strike would greatly reduce damage to the state; so the danger of crisis instability may lie less in the objective situation than in decision makers' incorrect beliefs.[74]

Most theorists who use the framework of anarchy take for granted not only people's instrumental beliefs but also the values that they place on outcomes. Although we realize that how the Prisoners' Dilemma is played—and indeed, whether it is a dilemma at all—is deeply influenced by the value each actor puts on the other's well-being,[75] scholars know little about the processes by which this evaluation is established and by which it can change. It is a central tenet in international politics that people value the security and well-being of their own state more than they do that of others. The self is defined as the national self. But this need not

[73] Snyder (fn. 29); Posen (fn. 44).

[74] For a further discussion, see Robert Jervis, *The Illogic of American Nuclear Strategy* (Ithaca, NY: Cornell University Press, 1984), 126-29; Jervis, "Psychological Aspects of Crisis Stability," in Jervis, *The Implications of the Nuclear Revolution* (Ithaca, NY: Cornell University Press, forthcoming); Richard Ned Lebow, *Nuclear Crisis Management: A Dangerous Illusion* (Ithaca, NY: Cornell University Press, 1987); Ashton Carter, John Steinbruner, and Charles Zraket, *Managing Nuclear Operations* (Washington, DC: Brookings Institution, 1987).

[75] Morton Deutsch, "Trust and Suspicion," *Journal of Conflict Resolution* 2 (December 1958), 265-79; Deutsch, "The Effect of Motivational Orientation upon Trust and Suspicion," *Human Relations* 13 (May 1960), 123-39. Similarly, rational choice analyses of politicians' behavior would yield very different results if it were assumed that the value being maximized was the individual's economic prospects rather than his power or votes. We would then expect politicians to cater to popular or constituency interests only to amass enough power that could be efficiently traded upon for pecuniary gain.

be true forever. It may be in the nature of human beings to put first their own well-being and that of others who carry their genes, but it is not inevitable that people will always care more about the fates of those on their side of a national border than they do about similar individuals on the other side.[76]

The issue is often put in terms of self-interest versus altruism, but it may be more useful to think of how the self is defined. Why should our attitudes toward others be based on their geographic location rather than on the values they hold?[77] Shared values can be one reason for national identity, but any individual can have more in common with many people in other countries than with many in his or her own. (Indeed, how Americans regard fellow-nationals seems less determined by their race and ethnicity than by the extent to which they share their values.)[78] Similarly, Alker and Sherman stress the importance of the "scope and depth of insecurity-relevant normative integration" in the international system as a determinant of foreign policy.[79]

The degree of value integration, and therefore the scope of the definition of the relevant self, may sometimes be larger than we assume. Statesmen usually like to present themselves as "hard-headed" and as caring almost exclusively about their own country. But policy is in fact often driven by motives that are hard for Realists to accommodate. For example, how can we explain the fact that the United States did not conquer Canada sometime in the past hundred years?[80] A Realist account would

[76] Some observers have attributed the relative lack of concern in the home country for the German hostages taken in Beirut in January 1987 to the weak German national identity. See James Markham, "West Germans Low-Key About Abductions," *New York Times*, January 19, 1987. Compare the reaction of Japan, a country some describe as "a huge tribal society," in a similar situation: Clyde Haberman, "Japan Outraged at Manila Abduction of Executive," *New York Times*, February 15, 1987.

[77] This question is an ancient one and can be traced back at least as far as Aristotle's advice that Alexander distinguish among his subjects according to whether they were Greeks or non-Greeks rather than according to their personal, individual characteristics; see *The Politics of Aristotle*, ed. and trans. by Ernest Barker (London: Oxford University Press, 1958), 388. For an attempt to use sociobiology to explain national loyalty, see R. Paul Shaw and Yuwa Wong, "Ethnic Mobilization and the Seeds of Warfare: An Evolutionary Perspective," *International Studies Quarterly* 31 (March 1987), 21-26.

[78] Milton Rokeach and Louis Mezei, "Race and Shared Belief in Social Choice," *Science* 151 (January 1966), 167-72; Rokeach, *Beliefs, Attitudes, and Values* (San Francisco: Jossey-Bass, 1968). Similarly, in the view of the Ottomans, "the community of true believers, . . . not the state, constitutes the basic Muslim policy, transcending all boundaries." See Thomas Naff, "The Ottoman Empire and the European States System," in Hedley Bull and Adam Watson, eds., *The Expansion of International Society* (Oxford: Clarendon Press, 1984), 143.

[79] Hayward Alker, Jr., and Frank Sherman, "Collective Security-Seeking Practices Since 1945," in Daniel Frei, ed., *Managing International Crises* (Beverly Hills, CA: Sage, 1982), 141-44. This essay draws on the work of Karl Deutsch, esp. Deutsch et al., *Political Community and the North Atlantic Area* (Princeton: Princeton University Press, 1957).

[80] William T.R. Fox, *A Continent Apart: The United States and Canada in World Politics* (Toronto: University of Toronto Press, 1985).

be strained, having to conjure up power-based disincentives that are hard to find. American cooperation is better explained by focusing on value integration in three senses. First, the use of force is negatively valued by most Americans unless the object is to remove a menace to American security or to establish a democratic regime. Second, many values are shared between citizens of Canada and of the United States. Even if conquest would have given the U.S. greater wealth or security, the fact that Canadian society resembles American means that most of what we want to see in a country is already in place in Canada. Third, largely because of the common values between the two countries, Americans prize the well-being of Canadians. Thus, using force against them would, by constituting an offense against Canadians, diminish the utilities of Americans as well.

Common and conflicting values also help to explain whom a state will offer to protect and whom it will oppose.[81] Karl Deutsch and his colleagues stress the role of the "compatibility of the main values held by the politically relevant strata of all participating units" in the formation of what they call security-communities—i.e., groups of nations among whom war is unthinkable.[82] This factor is also important in less drastic forms of cooperation. The American commitment to Europe—and still more, to Israel—cannot be accounted for apart from the common heritage and values that make Americans care about the fates of these peoples. Security considerations are insufficient to explain American ties, which would be deeply affected if Europe or Israel were to become fascist. Similarly, American opposition to communism in the third world is based not only on national security concerns, but also on an identification with human beings in other countries whom we do not want to live under tyranny.[83] It may then be quite wrong to see Realpolitik as the source of the decision to fight in Vietnam and to dismiss as cant the desire to save that country from misery. Is it an accident that the proponents of the war were wrong in their predictions of widespread "domino effects" of an American defeat, but were correct in their argument that millions of people in Indochina would suffer or die from a North Vietnamese victory?

The most far-reaching changes in international politics involve changes in national goals and values. Japan is now a much more suitable partner for cooperation than it was in the 1930s, and not only because ter-

[81] For a discussion of the role of ideology in alliances, see Ole Holsti, P. Terrence Hopmann, and John Sullivan, *Unity and Disintegration in International Alliances* (New York: Wiley, 1973), and Walt (fn. 48), 18-26.

[82] Deutsch (fn. 79), 46.

[83] For discussions of this argument, see Robert Jervis, *The Logic of Images in International Relations* (Princeton: Princeton University Press, 1970), 244-50; George Quester, *American Foreign Policy* (New York: Praeger, 1982).

ritorial expansion is neither possible nor economically necessary. Something has occurred that is more basic than changes in instrumental beliefs. Rabid nationalism and the drive to dominate have been transmuted. A Japanese nationalist of the 1930s who saw his country today would be horrified, as Mishima was. Because it took a cataclysm to produce such a change, this pattern does not provide an attractive route to a more cooperative world. But it does show both the importance and the mutability of values.

A less dramatic alteration may be a more hopeful model. In the late 1960s, West Germans came to accept the existence of the East German regime. The division of the country, while still deeply regretted, was no longer completely intolerable. With the territorial status quo generally accepted, tensions and the danger of war were reduced, and cooperation in the form of the Quadripartite and the Helsinki Agreements became possible.

In Realism, values are generally taken as unproblematic and constant; they are givens in game-theoretic studies. Thus, although analysis within the framework of anarchy can produce accurate results, it may make our treatment excessively static and distract us from important questions. How values are developed, maintained, and changed may be crucial to international politics, and may strongly influence the extent of cooperation; it is a matter, however, that cannot be explained within the anarchy framework.

Norms: Is Narrow Self-interest Enough? Hatred and Norms

Utilities can reflect passions as well as interests. Considerations of theoretical parsimony speak for ignoring greed, hatred, and envy, as well as morality and self-sacrifice; but that does not mean that such motives are absent in the world. Witness Lord Salisbury's analysis of Prussia's demands on France in 1870 (which, it can be argued, laid the foundations for World War I):

> Unless [the Germans] enjoy the pleasing sensation of witnessing the mortification of France, they will think that the objects of the war are only half-attained. . . . [I]t is revenge that they desire: not a strategic frontier, or the recovery of lost "brothers"—but terms of peace which shall drive the iron well into their enemy's soul.[84]

These are the sort of impulses that older Realist scholars from Thucydides to Morgenthau warn of. They lead to destructive policies and are dif-

[84] Lord Robert Salisbury, "Count Bismarck's Circular Letters to Foreign Courts, 1870," *The Quarterly Review* 129 (October 1870), 553. I am grateful to Marc Trachtenberg for pointing me to this article.

ficult to theorize about. Although Realism recognizes that cool, narrow self-interest may fail to guide behavior, it provides few grounds for predicting when it will fail or what behavior will follow if it does.

A related issue is the role of norms in furthering cooperation. In one sense, norms indicate behavior that is expected, standard, or normal. In another sense, they indicate behavior that is approved and valued positively.[85] But when do norms in the first sense become norms in the second sense? If they do not, can they really produce a great deal of cooperation? One of the themes of the anarchy literature, of course, is that narrow self-interest is sufficient under many conditions, but the evidence is not very strong and the conditions may be very restrictive.[86] For example, many scholars have argued that in the balance of power, stability can arise out of the interplay of the states' narrow national interests in survival and expansion. The results are not always peaceful, but they do maintain the system. Paul Schroeder's fascinating historical studies deny that this is an accurate picture of European balance-of-power politics, however. He argues that mediating structures of smaller states, a common understanding of international law, and a shared sense of the appropriate behavior were required if war was not to grow from a tool of statecraft to the dominating fact of international politics.[87]

In tracing much of the cause of World War I to the decline in Austria's position, Schroeder points to Britain: although it had a great stake in the

[85] Kratochwil and Ruggie argue that our standard methodology is inapproprite for verifying the existence of norms in the latter sense because pointing to instances in which norms are violated does not establish that they do not exist or are not important. See Friedrich Kratochwil and John Gerard Ruggie, "International Organization: A State of the Art on an Art of the State," *International Organization* 40 (Autumn 1986), 766-69. A more general treatment of norms along these lines is Friedrich Kratochwil, *Rules, Norms, and Decisions: On the Conditions of Practical and Legal Reasoning in International Relations and Domestic Affairs* (Cambridge: Cambridge University Press, forthcoming 1988).

[86] This question is raised, among other places, in Alker's analyses of how people play and think about Prisoners' Dilemma in the laboratory. See Hayward Alker, Jr., and Roger Hurwirtz, "Resolving Prisoner's Dilemmas" (unpub., M.I.T.); Alker, "Reflective Resolutions of Sequential Prisoner's Dilemmas," presented at the meeting of the Society for General Systems Research, May 30, 1985; and Alker, "From Quantity to Quality: A New Research Program on Resolving Sequential Prisoner's Dilemmas," presented at the 1985 meeting of the American Political Science Association. The incentives and settings of laboratory situations are so different from those operating in international politics, however, that it is far from clear that these experiments tell us much that can be directly transferred.

[87] See the following essays, all by Schroeder: "The Lost Intermediaries: The Impact of 1870 on the European System," *International History Review* 6 (February 1984), 1-27; "Containment Nineteenth-Century Style: How Russia Was Restrained," *South Atlantic Quarterly* 82 (Winter 1983), 1-18; "World War I as Galloping Gertie," *Journal of Modern History* 44 (September 1972), 319-45; "The Nineteenth-Century Balance of Power: Language and Theory," paper delivered at the 1977 meeting of the American Political Science Association; "The 19th-Century International System: Changes in the Structure," *World Politics* 39 (October 1986), 1-26. Also see Friedrich Kratochwil, "On the Notion of 'Interest' in International Relations," *International Organization* 36 (Winter 1982), 1-30.

status quo and in avoiding conflict, it never acted to limit or control dangerous shifts in the Balkans. Others also failed to maintain the necessary balance, although their behavior was logical to the extent that they were willing to tolerate a high risk of war. After showing how much Austria was weakened by Rumania's realignment in 1913, Schroeder stresses that despite its importance,

> no government addressed itself to the most obvious and critical question of all: how was this new, crucial development to be managed? How, that is, could it be harmonized with the overall European balance, incorporated into the prevailing international system without raising the already fearful strains to the point of explosion? No one thought about this problem or suggested doing anything about it.[88]

Without conscious management—without at least some states seeking the common interest of the members of the system—stability, peace, and the best interests of the individual states could not be maintained. A similar argument was made by Lord Salisbury, later to become Foreign Secretary, when he criticized the refusal of the British government to counteract the harsh terms Prussia imposed on defeated France:

> If [the leaders'] intention is . . . to draw all the profit they can from the arrangements of the greater international republic, and yet to bear no share of the cost and dangers of its government, we doubt not that they are preparing themselves for a severe condemnation from the English people. We only trust that they are not also preparing for England the national doom that always waits for the selfish and timid.[89]

A parallel point arises from one of Axelrod's computer simulations. He finds that stable cooperation is much more likely to occur when actors follow "metanorms" that call for punishing those who fail to enforce norms. In other words, actors must not only be prepared to punish those who defect, but also to act against those who fail to punish them.[90] In the same way, cooperation can be powerfully reinforced by injunctions to help others, or at least to limit the harm done to them. Thus "although prescriptive morality is not usually expressed in laws, in a number of Eu-

[88] Paul Schroeder, "Romania and the Great Powers before 1914," *Revue Roumaine D'Histoire* 14 (No. 1, 1975), 52-53. As Schroeder puts it in "World War I as Galloping Gertie": "Everyone wanted a payoff; no one wanted to pay" (fn. 87), 345.

[89] Salisbury (fn. 84), 556. During the Eastern Crisis of 1877, William Gladstone asked: "What is to be the consequence to civilisation and humanity, to public order, if British interests are to be the rule for British agents all over the world, and are to be for them the measure of right or wrong?" (Quoted in Seton-Watson, fn. 72, p. 69.) For a related general argument, see Armatya Sen, "Rational Fools: A Critique of the Behavior Foundations of Economic Theory," *Philosophy and Public Affairs* 6 (Summer 1977), 326-41.

[90] Robert Axelrod, "Modeling the Evolution of Norms," *American Political Science Review* 80 (December 1986), 1095-1111.

ropean countries and now also in some [American] states . . . , helping others in certain kinds of extreme need is required by law."[91]

Applying the same reasoning, Robert Trivers argues that evolution should select for people who reciprocate altruism. Emotions such as anger at those who do not reciprocate, and guilt over one's own cheating are also functional because they bolster reciprocity. Short-run calculations might lead people to ignore the transgressions of others (or to reply to them only mildly), or to shirk reparations when they have been caught cheating. But over the long run, such behavior is destructive. Disproportionate retribution for cheating, called up by anger, can induce future compliance: "it seems plausible . . . that the emotion of guilt has been selected for in humans partly in order to motivate the cheater to compensate his misdeed and . . . thus to prevent the rupture of reciprocal relationships." Because these emotions may sustain long-run cooperation more than calculation would,

> selection may favor distrusting those who perform altruistic acts without the emotional basis of generosity or guilt because the altruistic tendencies of such individuals may be less reliable in the future. One can imagine, for example, compensating for a misdeed without any emotional basis but with a calculating, self-serving motive. Such an individual should be distrusted because the calculating spirit that leads this subtle cheater now to compensate may in the future lead him to cheat when circumstances seem more advantageous.[92]

The issue is by no means closed. It is possible that narrow self-interest may be able to explain most instances of cooperation in the absence of binding authority. But we should note that in Gouldner's classic article, reciprocity is not only a common empirical pattern and a way out of a dilemma; it is also a moral imperative: "people *should* help those who have helped them."[93] Such a sense of obligation—if shared—may well

[91] Ervin Staub, *Positive Social Behavior and Morality*, I (New York: Academic Press, 1978), 3.

[92] Trivers, "The Evolution of Reciprocal Altruism," *Quarterly Journal of Biology* 46 (March 1971), 50-51. On p. 52, Trivers anticipates Axelrod's arguments about metanorms. A similar argument about deterrence is made by Dean Pruitt, "Some Relationships Between Interpersonal and International Conflict," in Axelrod et al. (fn. 3).

[93] Alvin Gouldner, "The Norm of Reciprocity," *American Sociological Review* 25 (April 1960), 169-71 (emphasis added); also see Larson (fn. 50), 20-22. For a nice summary of the normative hold of reciprocity, see Robert Cialdini, *Influence* (Glenview, IL: Scott, Foresman, 1985), 20-34, and especially the marvelous story on page 27. Staub notes that experiments indicate "that willingness to ask for help is reduced when people do not expect to have an opportunity to provide help in return" (fn. 91, p. 346). The reverse should have been found if rational calculation were the driving force. Also see Charles Kindleberger's review of *After Hegemony*, "Hierarchy vs. Inertial Cooperation," *International Organization* 40 (Autumn 1986), 844-46.

have great "practical value," as John Mackie notes.[94] Considerations of morality, fairness, and obligation are almost surely large parts of the explanation for the fact that individuals in society cooperate much more than the Prisoners' Dilemma would lead us to expect. Only economists behave as the theory says they should; others are likely to contribute to public goods, especially when they believe that fairness calls for them to do so.[95] Indeed, it is possible that morality provides the *only* way to reach many mutual cooperative outcomes. In part because of the tendency for people to be self-righteous and to see their own acts as cooperative and those of others as hostile, temptations and fears may produce mutually undesired outcomes as long as narrow self-interest is dominant. At a minimum, the feeling that one is morally obligated to reciprocate cooperation—and that others live under the same code—permits a wider range and scope for mutually beneficial exchanges. In fact, the actors may gain most when they do not regard the interaction as one of self-interested exchange at all. Even if this extreme is not approached (and it is not likely to be in international politics), without the power of at least some shared values, without some identification with the other, without norms that carry moral force, cooperation may be difficult to sustain.[96]

Conclusions

The queries and objections raised here are not all of the same type. Some are assertions that the anarchy framework leads us to concentrate

[94] John Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth, U.K.: Penguin, 1977), 119-20; see also Keohane (fn. 58), 126-27, and Keohane, "Reciprocity in International Relations," *International Organization* 40 (Winter 1986), 20-24.

[95] Gerald Marwell and Ruth Ames, "Economists Free Ride, Does Anyone Else?" *Journal of Public Economics* 15 (June 1981), 295-310. I am grateful to Joanne Gowa for referring me to this instructive article. Also see Charlan Nemeth, "A Critical Analysis of Research Utilizing the Prisoner's Dilemma Paradigm for the Study of Bargaining," in Leonard Berkowitz, ed., *Advances in Experimental Social Psychology*, VI (New York: Academic Press, 1972), 203-34; Daniel Kahneman, Jack Knetsch, and Richard Thaler, "Fairness and the Assumptions of Economics," *Journal of Business* 59 (October 1986), S285-300; Sen (fn. 89). Transcripts of the deliberations during the Cuban missile crisis reveal President Kennedy's concern with perceived fairness: McGeorge Bundy, transcriber, and James Blight, editor, "October 27, 1962: Transcripts of the meetings of the ExComm," *International Security* 12 (Winter 1987/88), 30-92.

[96] In his presidential address to the Public Choice Society, Dennis Mueller made a similar point: contrary to the logic of PD, in these situations

> most of us choose the cooperative strategy most of the time. Why? Because we were taught to do so. . . . One is almost embarrassed to make these observations were it not that so many of us who work with rational egoist models continually build our models on assumptions that ignore these truisms from psychology and everyday life.

Mueller, "Rational Egoism vs. Adaptive Egoism as a Fundamental Postulate for a Descriptive Theory of Human Behavior," *Public Choice* 51 (No. 1, 1986), 5-6.

on questions that are not central. That is true for the discussion of preferences. Even if my arguments about the importance of the preference orders and of changes in them are correct, the framework can still be applied, albeit at the cost of requiring us to treat as given what may be crucial and problematical. But if this objection is not fatal, it also applies more broadly to the use of game-theory models in general rather than to Prisoners' Dilemma in particular. Other problems complicate the picture. Basic concepts such as cooperation, offense, and power are not as straightforward as they appear in many analyses, and we may lose a great deal when we squeeze behavior into the categories the former two provide. Strong perceptual biases also undermine many of the central predictions derived from the PD. Although the national behavior that actually occurs is often consistent with the general Realist emphasis on conflict, the reasons are different from those stressed by this tradition. Game theory can accommodate both uncertainty and differences between the perceptions of the two sides,[97] but we must know what these perceptions are; they may be more important and difficult to understand than the resulting interaction. Problems that are even more fundamental arise if narrow self-interest is not the driving force behind national behavior: although game-theory models could be built around different premises, many of the Realist arguments about anarchy would be undermined.

While I have discussed the problems inherent in the anarchy framework, I have said little about its numerous strengths. To do so would be to recapitulate familiar arguments; but I do want to note that the framework is useful for reminding us that human action is often driven by the twin impulses of fear and temptation. It provides tools for analyzing how these impulses can be harnessed (if not tamed) in a way that leaves all parties better off. The concepts of anarchy and the security dilemma lead us to see that the international system not only permits conflict, but can create it: actors may refuse to cooperate with others, not so much because they seek the positive gains of exploitation, but because they fear that their own cooperative initiatives will be mistreated. For purposes of both analysis and prescription, the framework yields significant propositions on the conditions and strategies that increase the likelihood of cooperative behavior and outcomes. But its simplifications can be misleading, its assumptions require scrutiny, and it relegates many important questions to the background. Understanding and exploring both the uses of the approach and its limitations, without being overwhelmed by either, will lead to a better grasp of world politics.

[97] See, for example, John Harsany, "Bargaining in Ignorance of the Opponent's Utility Function," *Journal of Conflict Resolution* 6 (March 1962).