RATIONALITY AND DETERRENCE

By FRANK C. ZAGARE*

A rational deterrent cannot be based on irrational responses.
—*Richard Nixon*

MANY critics of deterrence theory rest their case against it on the inadequacy of the rational actor model.[1] Yet other students of interstate conflict, seemingly undeterred, continue to probe the subject of deterrence with models that explicitly postulate rational behavior.[2] Why?

One purpose of this essay is to suggest that since there is a critical, if frequently unappreciated, difference between the rational *actor* model and the assumption of rational *choice* sometimes taken to be synonymous with it, many of the criticisms are beside the point. A second is to show that some recent rational choice models of deterrence are, nonetheless,

* For helpful and insightful comments on an earlier version of this essay, I thank Bruce Bueno de Mesquita, Thomas Fogarty, Jacek Kugler, and Richard Ned Lebow.

[1] See, for instance, John H. Barton and Lawrence D. Weiler, eds., *International Arms Control: Issues and Agreements* (Stanford, CA: Stanford University Press, 1976); Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy* (New York: Columbia University Press, 1974); Patrick M. Morgan, *Deterrence: A Conceptual Analysis*, 2d ed. (Beverly Hills, CA: Sage, 1983); Robert Jervis, "Deterrence Theory Revisited," *World Politics* 31 (January 1979), 289-324.

[2] See, inter alia, Steven J. Brams, *Superpower Games* (New Haven, CT: Yale University Press, 1985); Steven J. Brams and D. Marc Kilgour, *Game Theory and National Security* (New York: Basil Blackwell, 1988); Bruce Bueno de Mesquita, *The War Trap* (New Haven, CT: Yale University Press, 1981); Bruce Bueno de Mesquita, "The War Trap Revisited," *American Political Science Review* 79 (March 1985), 156-73; Bruce Bueno de Mesquita and William H. Riker, "An Assessment of the Merits of Selective Nuclear Proliferation," *Journal of Conflict Resolution* 26 (June 1982), 283-306; Niall M. Fraser and Keith Hipel, *Conflict Analysis: Models and Resolution* (New York: North-Holland, 1984); David Gauthier, "Deterrence, Maximization, and Rationality," *Ethics* 94 (April 1984), 474-95; T. Clifton Morgan, "A Spatial Model of Crisis Bargaining," *International Studies Quarterly* 28 (December 1984), 407-26; James D. Morrow, "A Continuous-Outcome Expected Utility Theory of War," *Journal of Conflict Resolution* 27 (September 1985), 473-502; James D. Morrow, "A Spatial Theory of International Conflict," *American Political Science Review* 80 (December 1986), 1131-50. Robert Powell, "Crisis Bargaining, Escalation, and MAD," *American Political Science Review* 81 (September 1987), 717-35; Robert Powell, "Nuclear Brinkmanship with Two-Sided Incomplete Information," *American Political Science Review* 82 (March 1988), 156-78; R. Harrison Wagner, "Deterrence and Bargaining," *Journal of Conflict Resolution* 26 (June 1982), 329-58; R. Harrison Wagner, "The Theory of Games and the Problem of International Cooperation," *American Political Science Review* 77 (June 1983) 330-46; R. Harrison Wagner, "The Theory of Games and the Cuban Missile Crisis" (Paper presented at the Annual Meeting of the American Political Science Association, Chicago, IL, September 3-6, 1987); Frank C. Zagare, "Toward a Reformulation of the Theory of Mutual Deterrence," *International Studies Quarterly* 29 (June 1985), 155–69; and Frank C. Zagare, *The Dynamics of Deterrence* (Chicago: University of Chicago Press, 1987).

deficient in their application of the rationality postulate. Along the way, the theoretical implications of these points will be highlighted.

Procedural Rationality and the Case against the Rational Actor Model

In a well-known critique of contemporary deterrence theory, Patrick Morgan points out that "classic criticisms of deterrence theory turn on the charge that governments simply lack the necessary rationality to make it work, that they are particularly subject to irrationality in times of intense crisis or actual attack."[3] Implicit in this statement is a view of rationality which Simon calls *procedural* rationality.[4] In brief, theorists who subscribe to this view tend to equate rationality with omniscience. As described by Verba in an oft-cited article, a rational actor is one who makes a "cool and clearheaded ends-means calculation" after considering *all* possible courses of action and carefully weighing the pros and cons of each of them.[5] Obviously, such a decision requires that an actor have an accurate perception of the implications of all his options and a well-defined set of preferences concerning them. It also requires that he accurately assess the preferences of other relevant actors and their likely response to his tactical choices, that is, to his concessions or to his threats. In the view of the proceduralist, misperceptions—or other deficiencies of human cognition—and rational decision making are mutually exclusive. A rational agent—if one exists—will isolate himself from these hindrances. Moreover, he will also factor out of his decisional calculus other extraneous determinants stemming from psychological predispositions or emotional and affective deficiencies.[6]

In the context of national security issues, this view of rationality is articulated most explicitly by Allison in his classic study of the Cuban missile crisis.[7] Allison's critique of the rational actor model has had a profound impact on the study of international politics, and rightfully so.

[3] Morgan (fn. 1), 13.

[4] Herbert A. Simon, "From Substantive to Procedural Rationality," in S. J. Latsis, ed., *Method and Appraisal in Economics* (Cambridge: Cambridge University Press, 1976).

[5] Sidney Verba, "Assumptions of Rationality and Non-rationality in Models of the International System," in Klaus Knorr and Sidney Verba, eds., *The International System: Theoretical Essays* (Princeton, NJ: Princeton University Press, 1961), 95.

[6] Joseph de Rivera, *The Psychological Dimension of Foreign Policy* (Columbus, OH: Merrill, 1968); Richard Ned Lebow, *Between Peace and War: The Nature of International Crisis* (Baltimore, MD: The Johns Hopkins University Press, 1981); and John Steinbruner, "Beyond Rational Deterrence: The Struggle for New Conceptions," *World Politics* 28 (January 1976), 223-45.

[7] Graham T. Allison, *Essence of Decision: Explaining the Cuban Missile Crisis* (Boston: Little, Brown, 1971).

According to Mandel, "most scholars of the field no longer believe that states always use the 'billiard ball' rational actor approach; no longer treat psychological influences as random accidents or idiosyncratic deviations; and no longer assume that the most subjective aspects of international behavior are inherently unanalyzable."[8] Given the startling conceptual reorientation of the field since the heyday of political realism and the rational actor model,[9] the question that naturally arises is what role, if any, can be served by models that make explicit use of the rationality postulate to theorize about the nature of deterrence. To answer this, I distinguish first the procedural view of rationality implicit in most criticisms of the rational actor model from the instrumentalist definition of rationality generally associated with rational choice models.

Instrumental Rationality

In contrast to those who define rationality procedurally, those who define it *instrumentally* take a more limited view of this concept. Luce and Raiffa's is perhaps the clearest and most direct definition: a rational actor is one who, when confronted with "two alternatives which give rise to outcomes, . . . will choose the one which yields the more preferred outcome."[10]

This is not a deceptively simple definition. It *is* simple. There are no hidden assumptions lurking behind it. In fact, only two axioms, associated with the logical structure of an actor's preference function, are implicit in it. For an actor to be rational in the instrumentalist sense of Luce and Raiffa, he or she must have *connected* and *transitive* preferences over the set of available outcomes.

Connectivity simply means that an actor be able to make comparisons among the outcomes in the feasible set and evaluate them in a coherent way. For example, given a choice between two alternatives, a and b, a player with connected preferences will either prefer a to b, b to a, or be indifferent to both. Clearly, the behavior of an actor whose preferences

[8] Robert Mandel, "Psychological Approaches to International Relations," in Margaret G. Hermann, ed., *Political Psychology: Contemporary Problems and Issues* (San Francisco, CA: Jossey-Bass, 1986), 251.

[9] Of course, this paradigm shift of sorts is most properly traced to the pioneering work of Richard C. Snyder, H. W. Bruck, and Burton Sapin, eds., *Foreign Policy Decision-Making: An Approach to the Study of International Politics* (New York: Free Press, 1962). For this argument see James N. Rosenau, "The Premises and Promises of Decision-Making Analysis," in James C. Charlesworth, ed., *Contemporary Political Analysis* (New York: Free Press, 1967), and B. P. White, "Decision-making Analysis," in Trevor Taylor, ed., *Approaches and Theory in International Relations* (London: Longman, 1978).

[10] Duncan R. Luce and Howard Raiffa, *Games and Decisions: Introduction and Critical Survey* (New York: Wiley, 1957), 50.

are not connected is not amenable to analysis using a model or theory rooted in the notion of rational choice.

Transitivity implies the following: if an actor prefers alternative *a* to *b*, and *b* to *c*, then, if his preferences are transitive, he will also prefer *a* to *c*. If not, his preferences are logically incoherent and, as before, best analyzed outside a rational choice framework.

Surely these are minimal requirements for a definition of rationality. Without them, choice theory would be well-nigh impossible. In fact, *any* theory assuming purposeful action would also be impossible. More to the point, not only are these two assumptions implicit in all of the rational choice theories of deterrence so far developed, but I would submit that they are also implicit in what are sometimes seen to be incompatible theoretical constructs. For example, both of Allison's alternatives to the Rational Actor model (I), i.e., the Organizational Process model (II) and the Governmental Politics model (III) assume purposeful behavior. The organizations in Model II are said to pursue organizational goals, and the individuals in Model III are postulated to pursue political goals. For the same reasons given above, neither of these (conceptual) models could be applied without, at least implicitly, assuming that the preference functions of either the organizations or the individuals analyzed in them were connected and transitive. If these models are incompatible with one another, as Allison argues,[11] then it is simply because they postulate different units of analysis and make different assumptions about the goals these units pursue. They are not incompatible because they make fundamentally different assumptions about the underlying nature of the choices made by these units.

Nor are these assumptions either heroic or exceptional. While there may be instances of international decision makers suffering from mental illness, it is probably the case that most of them, including Hitler and Khomeini, have coherent—not laudable or even reasonable, but coherent—preference orders and are rational in the limited sense of the instrumentalist. Notice, however, that such a judgment does not rest upon an evaluation of the particulars of a decision maker's preferences. It merely assumes that these preferences are logically consistent, *whatever they may be*. Thus, while it is probably true that the context in which choices are framed may have a dramatic impact on a player's preference function,[12] rational choice models are not necessarily affected by this observation.

[11] Allison (fn. 7), 246.

[12] Amos Tversky and Daniel Kahnerman, "The Framing of Decisions and the Psychology of Choice," *Science*, 30 January 1981, 453-58.

The next question is where do these preferences come from, and how are they defined? For the instrumentalist, they are defined subjectively by each individual decision maker.[13] Thus, in contrast to the proceduralist, the instrumentalist does not presume to offer normative evaluations of an actor's preferences, however bizarre, reprehensible, or ill-founded they may be.[14] For instance, consider a leader who prefers systematic genocide to the benign neglect of a minority population. If his actions are consistent (or are perceived by the actor to be consistent) with this obviously repugnant ordering, he is rational by the definition of the instrumentalist. The reason is manifest: as a scientist, the instrumentalist is primarily interested in theory construction, not judging the ethical, strategic, political, or moral basis of an actor's motivation.[15] How best to understand Hitler's behavior? Simply by understanding his goals. In other words, preferences are given and either actual or optimal behavior deduced. The question of what preferences and/or perceptions an actor *should* have is considered not particularly relevant for developing explanatory or predictive theories of behavior.

Can an instrumentally rational actor have preferences rooted in incomplete, imperfect, or even erroneous information? Yes. Can variables which the proceduralist would reject as illegitimate influences on policy making have an impact on an actor's preferences? Yes. Are the distortions implied by the organizational context of policy making and the imperatives of the political process consistent with this notion of rationality? Yes. Can an individual whose vision is clouded by the pressures of time and stress in a crisis still be considered rational? In the limited sense of the instrumentalist, the answer, again, is yes.

Stating this in different terms: the instrumentalists' notion of rationality is not only quite restricted, but is also not necessarily inconsistent with conceptual models such as Lebow's based upon the notion of procedural rationality.[16] In fact, when used in some ways the instrumental-

[13] Daniel Ellsberg, "The Theory and Practice of Blackmail," lecture at the Lowell Institute, Boston, MA, March 10, 1959; reprinted in Oran R. Young, ed., *Bargaining: Formal Theories of Negotiation* (Urbana, IL: University of Illinois Press, 1975), 347.

[14] It may be true, as Robert Jervis writes in "Realism, Game Theory, and Cooperation," that "by taking preferences as given, we beg ... the most important question on how they are formed" (*World Politics* 40 [April 1988], 317-49, at 324). But this does not mean that other questions are unimportant. Indeed, as I argue below, meaningful scientific progress depends upon knowing both the sources of preferences and their strategic implications. Understanding how the world works should not always be depicted as a zero-sum game between competing groups of theorists, or as an either/or choice about research foci and methods.

[15] This is not to suggest that normative questions are outside the scope of legitimate inquiry. It is simply to say that, in general, instrumentalists have used this concept for purposes that are different from, though not necessarily inconsistent with, more traditional normative concerns.

[16] Lebow (fn. 6).

ists' definition is nothing more than a convenient tautology.[17] Instrumentalists—and proceduralists as well—use it not because it is the "correct" way of defining this term, but because the assumption of instrumental rationality is useful for constructing theories of rational or psychological choice.

The point, however, is that rational choice models, as opposed to the rational (or unitary) actor model in international politics, should not be rejected, a priori, by those who define rationality procedurally. As used by the instrumentalist, this term does not connote superhuman calculating ability, omniscience, or an Olympian view of the world, as some proceduralists have concluded.[18] The individual decision makers analyzed by rational choice theorists can be, at one and the same time, rational in the limited instrumental sense, and irrational in the sense of the proceduralist. Thus, to the extent that subjective interpretations of the world are built into the models of the instrumentalist,[19] such models could also be used to describe the behavior of decision makers suffering from cognitive closure, selective perceptions, misinformation and so on.[20] There is no necessary inconsistency between these two seemingly divergent notions of rationality.

Some Implications of the Assumption of Instrumental Rationality

Before turning to an assessment of some recent rational choice models of deterrence, it will be useful first to explore one development in noncooperative game theory which has deepened, considerably, our knowledge of instrumental rationality, namely, Selten's concept of a *perfect equilibrium*.[21] Until the development of this concept, most game theorists were of the opinion that the only universally defensible rationality requirement for players in a noncooperative game was that their strategies form a *Nash equilibrium* pair (outcome).[22] What Selten did, by way of

[17] E.g., Samuelson's theory of revealed preferences. Paul A. Samuelson, "A Note on the Pure Theory of Consumer's Behaviour," *Economica* 5 (February 1938), 61-71.

[18] E.g., Glenn H. Snyder and Paul Diesing, *Conflict Among Nations: Bargaining, Decision Making and System Structure in International Crises* (Princeton, NJ: Princeton University Press, 1977), chap. 5.

[19] See, for example, Bueno de Mesquita's expected utility model developed in *The War Trap* (fn. 2) and "The War Trap Revisited" (fn. 2), which postulates decision makers who, by way of their risk functions, distort objective reality.

[20] See Arthur A. Stein, "When Misperception Matters," *World Politics* 34 (July 1982), 505-26, for an example of such a merger.

[21] Reinhard Selten, "A Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory* 4, No. 1 (1975), 25-55.

[22] John Nash, "Non-cooperative Games," *Annals of Mathematics* 54 (September 1951), 286-95.

counterexample, was to demonstrate that only perfect Nash equilibria could be defended in terms of the rationality principle.

Backtracking for a moment, let us define a few terms: a *noncooperative* game is one in which the players are either unable to communicate with one another or, because of the lack of an enforcement mechanism, to commit themselves to any particular strategy. It is easy to see why noncooperative game theory has held a particular attraction for theorists of interstate conflict. In the absence of an overarching authority in the interstate system to enforce commitments or agreements, great power politics clearly meets the definitional requirements of a noncooperative game.

At the heart of the theory of noncooperative games lies Nash's equilibrium concept. A Nash equilibrium is an outcome from which neither player can gain, immediately, by switching to another strategy (outcome). The reason for the centrality of this concept in noncooperative game theory is apparent: since Nash equilibria are associated with instrumentally rational choices by both players, they are, in essence, self-enforcing outcomes. Hence, in an environment which lacks an enforcement agent, only Nash equilibria are rational and one would expect that only these outcomes would be selected by rational players. All outcomes which do not meet the Nash criteria involve irrational behavior since at least one player has an incentive to switch his strategy to induce a more preferred outcome.[23]

To illustrate these concepts, consider Figure 1, a game originally developed by Harsanyi.[24] In this game, there are two players, A (row) and B (column). Each player has two strategies: to either cooperate (C) or not to cooperate (D) with the other. These choices give rise to four outcomes. In Figure 1, the outcomes are represented by an ordered pair representing the payoff to A and B, respectively. The outcomes are ranked from best to worst, with the highest ranked outcome indicated by the largest integer, the next-highest outcome by the next-largest integer, and so on.

There are two Nash equilibria in Figure 1, each indicated by an asterisk. Outcome (2,2) is in equilibrium since each player will do worse by switching, unilaterally, to another strategy and inducing another out-

[23] This is why a game theorist would object to Karl W. Deutsch's suggestion that *each* player in a Chicken game choose his cooperative (C) strategy. Since the resulting outcome (3,3) is not an equilibrium outcome, it is inconsistent with the instrumentalists' definition of rationality. See Deutsch, *The Analysis of International Relations*, 3d ed. (Englewood Cliffs, NJ: Prentice-Hall, 1988), 149-50.

[24] John C. Harsanyi, "Advances in Understanding Rational Behavior," in R. E. Butts and J. Hintikka, eds., *Foundational Problems in the Special Sciences* (Dordrecht, Holland: D. Reidel, 1977).

		B: Cooperate (C)	B: Not Cooperate (D)
A	Cooperate (C)	(1,3)*	(1,3)
	Not Cooperate (D)	(0,0)	(2,2)*

FIGURE 1
HARSANYI'S GAME

* = Nash equilibrium

come. Specifically, were A to switch from his (D) strategy, which supports (2,2), to his (C) strategy, which would induce outcome (1,3), his payoff would go from "2"—his best—to "1"—his next best. And if B were to switch to her (C) strategy, her payoff would go from "2"—her next best—to "0"—her worst. Thus, since neither player benefits by switching unilaterally to another strategy, (2,2) is a Nash equilibrium. For similar reasons, the (1,3) outcome marked with an asterisk is also a Nash equilibrium. By contrast, neither of the other outcomes is stable in the sense of Nash.

Although there are two Nash equilibria in this game, they are not equally defensible as outcomes rational players would select. Specifically, since (2,2) is the product of a dominant strategy[25] of B, and the best response to this strategy by A, compelling reasons exist to consider this outcome, and not the other equilibrium, (1,3), as the solution to this game.

One might object, however, by arguing that B could do better. Could not B threaten to choose her (C) strategy if A selects (D), thereby inducing A to choose (C) and bringing about B's best outcome? No, because the equilibrium at (1,3) is not perfect, that is, because it involves both "irrational behavior and irrational expectations by the players about each other's behavior."[26] To see why, consider the game-tree representation of this game given in Figure 2.

When the game of Figure 1 is represented in extensive or game-tree

[25] A dominant strategy provides at least as good an outcome for a player as any other, no matter what strategy the other player(s) in the game select. For a further discussion of this and related concepts, see Frank C. Zagare, *Game Theory: Concepts and Applications* (Beverly Hills, CA: Sage, 1984).

[26] Harsanyi (fn. 24), 332.

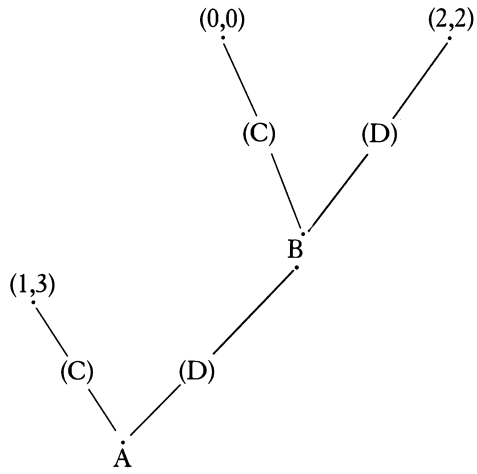

FIGURE 2
GAME-TREE REPRESENTATION OF HARSANYI'S GAME

form, it is easy to see why (1,3) can be eliminated as an outcome (instrumentally) rational players would select. Consider the calculus of A at the first node of this tree. A can either select (C) and induce (1,3), his second-best outcome, or (D), which can result in either his best or worst outcome. Clearly, he should choose (C) if he expects B to also select (C) since the choice of (D) would then result in his worst outcome. Conversely, he should select (D) if he expects B to select (D) since this induces his best outcome at (2,2). The question is: what should A expect B to do?

If A assumes that B is rational, he should expect B to select (D), since (D) is at least as good for B as (C), and sometimes better, no matter what A does. To expect B to carry out her threat to choose (C) if A chooses (D) is to assume that B is irrational. And for B to expect A to select (C) is to assume that B harbors irrational expectations about A. This is why (1,3) is not perfect, that is, an equilibrium rational players would select.

Still, one might object that B could "irrevocably commit" herself to (C), leaving A no choice. Might not such a commitment change the answer? As will be seen, the response to this question has important implications for the logical coherency of what Achen and Snidal call "rational deterrence theory."[27] I shall provide an answer in due course.

[27] Christopher H. Achen and Duncan Snidal, "Rational Deterrence Theory and Comparative Case Studies," *World Politics* 41 (January 1989), 143-69.

Rationality and Deterrence

Previously I suggested that a widespread misunderstanding of the instrumentalists' view of rationality exists among a large number of political scientists. As a consequence of this misunderstanding, a totally counterproductive and avoidable schism between rational choice theorists and scholars who define rationality procedurally has prevented a meaningful dialogue between them.[28] To be sure, the proceduralist's view of rationality has predisposed him to develop a brand of theory that looks and feels very different from the theory produced by scholars who assume only instrumental rationality. The wise instrumentalist, however, would be ill-advised to ignore this strand of research since rational choice models are not only potentially consistent with models or theories stressing individual-level variables, but also presuppose them. In other words, one can interpret the work of micro-level theorists as exploring the causal field of cognitions and affectations culminating in preference functions and the perception thereof. Similarly, one can interpret the work of the rational choice theorist as exploring the strategic consequences of various sets of real or perceived preference vectors. Putting this in still another way, to fully flesh out a rational choice model, a theory of preference formation is required. And, as many proceduralists implicitly acknowledge, to completely understand the consequences of perceptions and misperceptions, a theory of strategic interaction, like game theory, is needed. Thus, not only are these two definitions of rationality consistent with one another, but the theories that flow from them are potentially synergistic.[29]

Merely recognizing this potential compatibility, however, will not make the task of theoretical integration easier. One reason is that, as Jervis points out, a parsimonious and coherent theory of preference formation based on micro-level variables does not yet exist.[30] An equally important reason, though, is that deterrence theory itself is not yet fully and consistently specified.[31] I will now explain why this, unfortunately, is so.

[28] The irony of this misunderstanding is hard to miss.

[29] George W. Downs, "The Rational Deterrence Debate," *World Politics* 41 (January 1989), 225-37.

[30] Robert Jervis, "Perceiving and Coping with Threat," in Robert Jervis, Richard Ned Lebow, and Janice Gross Stein, eds., *Psychology and Deterrence* (Baltimore, MD: The Johns Hopkins University Press, 1985), 33.

[31] Christopher H. Achen, "A Darwinian View of Deterrence," in Jacek Kugler and Frank C. Zagare, eds., *Exploring the Stability of Deterrence* (Denver: University of Denver School of International Studies, 1987); Zagare (fn. 2, 1987). This is why Robert Jervis, "Introduction:

As almost everyone who has written about deterrence has pointed out, the underlying ideas behind this concept are not new. But deterrence as an important strategic concept only came into its own after the bombing of Hiroshima and Nagasaki in 1945. The reason for this was straightforward. As Brodie wrote soon afterward, up to this point "the chief purpose of our military establishment has been to win wars. From now on its chief purpose must be to avert them."[32]

Although it took time for some strategists to appreciate the wisdom of Brodie's initial insight, by the mid-1950s the concept of nuclear deterrence had emerged as the first among equals in the pantheon of U.S. strategic principles. In the earliest days of the postwar period, before the era of nuclear plenty, the containment policy of the Truman administration was, by necessity, augmented by a massive buildup of conventional weapons. But the "New Look" policy of the Eisenhower administration, stressing as it did nuclear weapons delivering "more bang for the buck," changed all of this. Now the United States would deter Soviet expansionism by threatening massive retaliation if and when the status quo was violated. The huge absolute costs associated first with atomic and then thermonuclear weapons would guarantee a stable international order. And when the Soviets themselves developed a meaningful strategic nuclear capability in the early 1960s, the doctrine of Massive Retaliation evolved into the notion of Mutual Assured Destruction.

Again, the supposition was that the massive costs associated with nuclear war would deter both superpowers. The absence of a superpower war since 1945 has merely reinforced the widely shared view that nuclear deterrence constitutes an unusually robust and stable relationship. In fact, to this day, the majority of Western strategic thinkers hold that the existence of the U.S. nuclear deterrent is uniquely responsible for the stability of the international system since 1945.

The problem with this explanation, however, is that it is logically flawed. As critics of the New Look policy were quick to point out,[33] while the notion that an aggressor can be deterred by the threat of massive costs is a reasonable hypothesis, provided that the costs of inflicting

Approach and Assumptions," in Jervis, Lebow, and Stein (fn. 30), 6, correctly, finds that "many events present unexplained puzzles for standard deterrence theory." For an attempt to explain some of these inconsistencies, see Jacek Kugler and Frank C. Zagare, "The Long-term Stability of Deterrence," *International Interactions* 15, Nos. 3 and 4 (1989), 253-75.

[32] Bernard Brodie, ed., *The Absolute Weapon: Atomic Power and World Order* (New York: Harcourt Brace, 1946), 76.

[33] See, for example, William Kaufmann, "The Requirements of Deterrence," in William Kaufmann, ed., *Military Policy and National Security* (Princeton, NJ: Princeton University Press, 1956).

the punishment flow only one way, once this relationship is symmetric, the logic of the argument breaks down.

To see this, consider now Figure 3 which summarizes, albeit at the expense of a great deal of conceptual simplification, the problem of mutual deterrence in a nuclear relationship. Underlying this figure are the following assumptions. First, each state has two broad strategic choices, to either cooperate (C) with the other by supporting the status quo, or not to cooperate (D) by seeking to overturn it. These choices give rise to four, equally broad outcomes: if each state cooperates, the status quo reigns; if one cooperates and the other does not, the state that does not gains an advantage. And if neither cooperates, conflict (read nuclear war) is implied.

Figure 3 further reflects the assumption that each side prefers an advantage to the status quo, but that each prefers the status quo to the other gaining an advantage. (As before, the outcomes are ranked from best to worst, with the higher-ranked outcomes represented by larger integers.) Finally, it mirrors the argument of many deterrence theorists that, in the nuclear age, the costs of conflict are so high that this outcome is worst for each state. The derivative game, called "Chicken," is the standard metaphor for deterrence in the nuclear age.[34]

		Nation B	
		Cooperate (C)	Not Cooperate (D)
Nation A	Cooperate (C)	Status quo (3,3)	Advantage to B (2,4)
	Not Cooperate (D)	Advantage to A (4,2)	Conflict (1,1)

Figure 3
The Paradox of Mutual Deterrence (Chicken)

[34] See, inter alia, Jervis (fn. 1), 291; Anatol Rapoport, *Strategy and Conscience* (New York: Harper and Row, 1964), 116; Raymond F. Hopkins and Richard W. Mansbach, *Structure and Process in International Politics* (New York: Harper and Row, 1973), 368-69; Steven J. Brams, *Game Theory and Politics* (New York: Free Press, 1975); Brams (fn. 2); Brams and Kilgour (fn. 2); Thomas C. Schelling, *The Strategy of Conflict* (Cambridge: Harvard University Press, 1960); and Thomas C. Schelling, *Arms and Influence* (New Haven, CT: Yale University Press, 1966).

If the game of Chicken accurately reflects the structural and psychological conditions of nuclear deterrence, then the problem with the theory is clear: assuming instrumental rationality, deterrence should not work. Given that each player initially chooses to cooperate, i.e., chooses (C), each player also has an incentive not to cooperate, i.e., switch to (D). Neither is deterred since neither should fear retaliation by a *rational* opponent.

To see why, suppose that one state, say B, upsets the status quo by switching from its (C) strategy to its (D) strategy, thereby gaining a momentary advantage. At this point the present outcome of the game is (2,4). Now, what is the rational response of State A? A has two choices: to stay at (2,4) by continuing to cooperate, or to move to (1,1) by defecting and executing its deterrent threat. Since A (by assumption) prefers (2,4) to (1,1), it should, if it is instrumentally rational, "chicken out" and accept B's transgression. And if A can be expected not to retaliate, then B should not hesitate to upset the status quo.

By symmetry, the same argument would apply were A to move from the status quo first. Consequently, if rational actors with the ability to inflict enormous costs on each other in a second strike are assumed, then the deterrence relationship is unstable. Furthermore, the apparent faith of many strategic theorists in the stabilizing consequences of nuclear weapons would seem to be unfounded. It appears, therefore, that there is an inherent contradiction between the precepts of instrumental rationality and the intuition of many strategists about the stabilizing impact of nuclear weapons. This is one reason why Jervis asserts that "a rational strategy for the employment of nuclear weapons is a contradiction in terms."[35]

Analytically, one is left with two options: (1) reject deterrence theory as logically incoherent, as does Achen,[36] or (2) rescue the theory by solving the paradox and demonstrating that deterrence can in fact be rational.[37] Starting with Ellsberg[38] and Schelling,[39] and continuing to more recent modeling efforts, there has been a long pedigree of attempts to resolve the paradox and demonstrate that deterrence can be an instrumentally rational policy under the conditions specified above. Have these

[35] Robert Jervis, *The Illogic of American Nuclear Strategy* (Ithaca, NY: Cornell University Press, 1984), 19.

[36] Achen (fn. 31).

[37] There is actually a third option. For their part, micro-level theorists have attempted to recast the theory by placing it on a firmer psychological foundation. This, in my opinion, is constructive, but not necessarily at odds with the second alternative.

[38] Ellsberg (fn. 13).

[39] Schelling (fn. 34, 1960, 1966).

attempts to reconcile the observed stability of the postwar era with the theoretical instability exhibited by the status quo outcome in the game of Chicken been successful? No. Though many of these efforts are rich and suggestive, they are also inadequate. Next, I explain why some recent attempts to eliminate the paradox have failed.[40]

Prominent Attempts to Resolve the Paradox of Deterrence

Underlying many of the proposals to resolve the paradox of mutual deterrence is the supposition that a player can somehow commit himself to a particular strategy. In this regard, one should distinguish two kinds of commitments. The first—a pre-commitment—is made to a strategy which is executed *before* the other player has selected his; the second—a post-commitment?—is made to a strategy which is executed *after* the other player has made a strategy choice. In the former case, it is easy to see that if a player is able to preempt the choice of the other in Chicken, the preempting player will win. By choosing (D) first, a preempting player can force his opponent to choose between his worst and next-worst outcomes. Numerous tactics for making pre-commitments have been advanced by Schelling, Ellsberg, Snyder, Jervis, Kahn, and others.[41] If nothing else, each of these efforts underscores the paradox of mutual deterrence because each is predicated upon the premise that the player choosing second, being instrumentally rational, will concede.

It is the second type of commitment, however, that is usually called upon to explain the stability of nuclear deterrence. It is easy to understand why such a contingent commitment, if believed, will deter the other side from selecting (D) first. If the other believes his opponent will respond with certainty[42] to a choice of (D) with (D) himself, his alternatives are reduced to choosing (C)—which would induce his next-best outcome—or (D)—which would induce his worst outcome. Given this choice, the other player would not rationally preempt his opponent. And if both players can commit themselves to a tit-for-tat strategy of responding to a (D) strategy with a (D) strategy, then mutual deterrence is established.

[40] For a critique of the logical foundations of the earlier theoretical literature, see Zagare (fn. 2, 1987), chap. 1.

[41] Schelling (fn. 34, 1960, 1966); Ellsberg (fn. 13); Glenn H. Snyder, "Crisis Bargaining," in Charles F. Hermann, ed., *International Crises: Insights From Behavioral Research* (New York: Free Press, 1972); Robert Jervis, "Bargaining and Bargaining Tactics," in J. Roland Pennock and John W. Chapman, eds., *Coercion*, Nomos XIV, Yearbook of the American Society for Political and Legal Philosophy (Chicago: Aldine-Atherton, 1972); Herman Kahn, *On Thermonuclear War* (Princeton, NJ: Princeton University Press, 1960).

[42] Or, as many have noted, with a "sufficiently" high probability.

But, and this is the crux of the matter, can a rational player make such a commitment given the costs of carrying out the retaliatory threat? Gauthier argues that he can.[43] To support his contention, Gauthier constructs an expected utility model of deterrence wherein the costs of nuclear retaliation exceed the costs of capitulation, thereby making retaliation irrational or incredible. Next, he queries whether a utility maximizer can commit himself to retaliate even when retaliation is inherently incredible. He can, Gauthier claims, but only if a commitment to retaliation provides the actor with a higher expected payoff than noncommitment. Then, not surprisingly, he shows that under certain circumstances such a commitment can indeed be rational, although Gauthier is fully cognizant that the absence of these same conditions renders a retaliatory strategy irrational and deterrence unstable.

Gauthier does not argue that it is rational to form the intention to retaliate if and only if it is utility-maximizing to execute it. Rather, he argues that it is rational to execute the intention if and only if it is utility-maximizing to form it. From this it follows that since it may be rational to form the intention to execute the threat, it may also be rational to carry it out. In fact, Gauthier—with laudable consistency—asserts that if the intention has been formed, and deterrence fails, then a rational agent who intends to retaliate should do so because acting upon this intention is part of the behavior required of an expected utility maximizer.

Gauthier's contention that a utility-maximizing agent may rationally retaliate in the case of a deterrence failure sets his argument apart from a similar one advanced by Brams and Kilgour who recognize that, if deterrence fails, it will always be irrational to retaliate.[44] Other than this, however, Brams and Kilgour's work is similar in spirit, though not in detail, to Gauthier's. They begin with an underlying model of deterrence based upon the structure of Chicken, but produce a qualitatively different game by permitting the players to commit themselves to quantitative levels of preemption and retaliation strategies. Given these assumptions, they show that a deterrence equilibrium can emerge in the game they construct, along with two other equilibria, each of which involves preemption by one player. It is precisely for this reason that Brams and Kilgour conclude that deterrence can constitute a rational and stable relationship.

There are, however, two problems with accepting this conclusion. First, since the stability of deterrence in either Gauthier's or Brams and Kilgour's model depends upon each player's commitment to the (pres-

[43] Gauthier (fn. 2).

[44] Brams and Kilgour (fn. 2).

ently) irrational should deterrence fail, each resolution of the paradox fails to satisfy Selten's perfectness criterion. This raises the question of why either player should believe that his opponent is actually committed to retaliation. If it is (at least momentarily) irrational to retaliate in the event of a breakdown of deterrence, one can only be deterred if one believes that one's opponent is, or will be, instrumentally (and procedurally) irrational after an attack. But if a player is deterred because he believes his opponent will retaliate irrationally, how can he simultaneously defend a policy predicated on the assumption that the same opponent, in being deterred, will be perfectly rational? Or put differently, why manipulate the cost function of an opponent by threatening massive levels of destruction when the overall stability of the relationship can only be explained if each player, at some point, is assumed to be willing to completely disregard these costs and act irrationally? In still other words, a theory of deterrence that explicitly rejects the perfectness criterion is a contradiction in terms. In effect it explains deterrence stability by assuming, concurrently, that a player is rational (when he is deterred) and irrational (when he is deterring his opponent). Such an intellectual sleight of hand is known as having it both ways.

Also problematic is the assumption that players are able to commit themselves to retaliate. Recall, however, that such commitments are not permitted in noncooperative game theory and that the paradox of deterrence rests upon the inability of the players to commit themselves to any particular strategy. Thus, these two resolutions solve the paradox by assuming away the very source of the difficulty.

At this point one might respond "so what?" In the real world the predefined rules of game theory are not inviolate. Granted. Let us accept this premise for a moment. Let us assume a world in which such commitments are possible, or to use Schelling's phrase, a world where " 'cross my heart' is universally recognized as absolutely binding."[45]

It is clear that in such a world there would be no security dilemma. To see why, assume that the players are able to commit themselves to a particular strategy in the deterrence game of Figure 3 (Chicken). Given this assumption, each player could simply agree to choose (C) and maintain the status quo. In so doing, each player would implicitly agree to forgo the individual incentive he has to upset the status quo. There would be no problem with this agreement since, by assumption, it is strictly enforceable. Thus, if strategy commitments are permitted, there is no paradox.

[45] Schelling (fn. 4, 1960), 26.

Still, many deterrence theorists would instinctively reject any proposal to stabilize a deterrence relationship which depended upon an opponent forgoing his interests. Such proposals would be rejected as utopian or idealistic. A fundamental principle of political realism, of which deterrence theory is an intellectual descendant, is that states are self-interested power maximizers. But it is precisely this premise that Gauthier and Brams and Kilgour use to explain deterrence stability. Putting this another way, it is simply inconsistent to assume that states will *not* forgo individual benefits and, at the same time, hold that the stability of deterrence rests upon the willingness of each state to carry out a threat which is not only instrumentally irrational, but also incompatible with the proposition that states seek, first, self-preservation and second, power maximization.

In sum, either states can make commitments or they cannot. If they cannot, then the arguments of Gauthier and Brams and Kilgour are not germane. And if commitments are permitted, then one cannot reject, a priori, other proposed solutions to international security as hopelessly utopian and remain logically consistent at the same time.[46]

Recognizing this, Powell has advanced a different line of reasoning to explain the rationality of deterrence.[47] Like others before him, Powell begins with an underlying Chicken model which is transformed, via assumption, into a sequential game in which the players must decide whether to escalate or (by submitting) not escalate, or to attack. If a player chooses not to escalate or attack, the game ends. If escalation is chosen, the second player is faced with similar choices. The four possible outcomes are the same as those in Chicken. If the player choosing first does not escalate or attack, the status quo prevails. If one player escalates and the other does not, the escalating player wins. If either player attacks, the game ends in disaster. And if both players escalate, the game continues until one player submits or until the game gets out of control and culminates in disaster. Powell assumes that by choosing to escalate, a player unleashes an autonomous risk, beyond his control, of disaster. Thus, his model captures well the spirit of the view of a nuclear crisis as a "competition in risk taking."[48]

[46] Gauthier (fn. 2), 494, to his credit, maintains logical consistency by admitting other possible commitment strategies. "Rational nations," he writes, "recognizing the need to seek peace and follow it given the costs of war, can unilaterally renounce the first use of nuclear weapons and thereby end all strike policies." But if this prescription strikes the reader as hopelessly naive, then so should the prescriptions of deterrence theory. Each rests upon self-abnegating choices. Brams and Kilgour assume the same when they allow each player to commit, albeit probabilistically, to a level of nonpreemption.

[47] Powell (fn. 2, 1987).

[48] Powell's model thereby provides a formalization of Schelling's "strategy-that-leaves-something-to-chance" (fn. 34, 1960).

Given these assumptions, Powell shows that the existence of a crisis equilibrium, meaning a stable outcome that arises after a challenge by one of the players and resistance by the other, depends on incomplete information, that is, each player's lack of information about the (cardinal) utility function of his opponent.[49] Moreover, concerning the purposes of this essay, he demonstrates that, under certain conditions, no challenge will be made and, hence, deterrence is stable. This suggests that even if each of two states knows the other prefers capitulation to disaster, neither would issue a challenge, provided that the resolve of each player to resist a challenge by the other passes a certain threshold. Interestingly, Powell's model reveals that when deterrence breaks down the connection between a state's resolve and its emergence as the victor in a crisis does not always depend upon a greater willingness to risk war.

One might think that this puts the issue to rest. Even in a world in which mutual defection is the worst of all possible outcomes, a rational player can choose *not* to challenge the other because each can obtain a higher expected payoff by following this course. Moreover, while the threat to unleash disaster is not seen to be credible, the threat to escalate and risk war satisfies the rationality, and hence, credibility (i.e., perfectness) criterion. Thus, Powell's model would seem to explain the stability of deterrence in the nuclear age. Each superpower may have been deterred from attacking the other because it feared that the other would, by resisting, unleash a process that would escalate and get out of control.

The problem with this conclusion, however, lies in the nature of the assumptions necessary to generate it. First, note that Powell assumes the players know the ordinal, though not the cardinal, valuation of each other's payoffs. Also note that the deductions rest upon the supposition that the choice of "attack" by one player *always* results in mutual disaster. At the highest rung of the escalation ladder, therefore, each player's threat to retaliate is afforded perfect credibility, even though it is instrumentally irrational to carry out. Thus it is not surprising that none of the equilibrium strategies identified by Powell involve a direct attack by one player against the other. Such a possibility is eliminated by assumption. But it is precisely this assumption that has been questioned by those who argue that deterrence could fail should a window of vulnerability open and tempt one side to launch a limited first strike, thereby creating and then exploiting a nuclear asymmetry.[50] Or put in another way, at the strategic level, Powell's model postulates rather than derives stability.

[49] When information is complete, deterrence is never stable. The player with the highest "effective" resolve simply escalates and wins. A similar conclusion is found in Zagare (fn. 2, 1987), 53-54.

[50] Colin S. Gray, "Nuclear Strategy: The Case for a Theory of Victory," *International*

Leaving this problem aside for the moment, the next question is whether or not deterrence can emerge at some lower level during a crisis interaction in which at least one side has challenged the status quo and the other has resisted the incursion. Or put differently, given the overall stability at the highest level of general strategic deterrence, why doesn't one side or the other simply escalate to the penultimate stage of the game since, by assumption, each player is deterred in the next and last stage of the game?[51] The answer suggested by Powell's model is that such an escalatory process would not occur—under specified conditions—because of the fear on each side that the other might do something to cause the process to get out of control. But why should either player fear that his opponent would unleash a process which would lead down the slippery slope toward a general nuclear war? Given the payoff assumptions noted above, if an opponent is (instrumentally) rational, such a fear would appear to be unfounded. As Maxwell correctly points out, "if the supposition that neither side believes the other would deliberately initiate nuclear war is accepted . . . neither side would have any reason to believe that there was a 'danger of things getting out of hand.' "[52]

Powell's answer is that such a process would not be selected by either player but by a third player, Nature, who, without a stake in the game, would impose the sanction probabilistically. Thus, in this model, crisis stability depends not only upon the stipulation of an irrational response at the highest rung of the escalation ladder, but also upon the assumption that, at lower levels, the irrational will occur with some positive probability. This is precisely why Achen argues that "far from leaning too heavily on rational choice postulates, 'rational deterrence theory' necessarily assumes that nations are not always self-interestedly rational."[53] The "strategy-that-leaves-something-to-chance" can, therefore, explain deterrence stability only if the rationality principle is stood on its head.[54]

Security 4 (Summer 1979), 54-87; Paul H. Nitze, "Deterring Our Deterrent," *Foreign Policy* 25 (Winter 1976-77), 195-210; Albert Wohlstetter, "The Delicate Balance of Terror," *Foreign Affairs* 37 (January 1959), 211-35.

[51] Why, for instance, don't the Soviets simply invade western Europe, given that each side's strategic arsenal is mutually deterred?

[52] Stephen Maxwell, *Rationality in Deterrence*, Adelphi Paper No. 50 (London: Institute for Strategic Studies, 1968).

[53] Achen (fn. 31), 92.

[54] Powell (fn. 2, 1987) 725, admits as much when he writes "One might object that requiring the states' strategies to be sequentially rational and then relying on Nature to impose the irrational sanction does not really solve the credibility problem. I agree with this criticism." Powell goes on to note, however, that "it is important to realize that this is not so much a criticism of the model as it is a fundamental criticism of the way that the strategy-that-leaves-something-to-chance has attempted to overcome the credibility problem. The model only exposes this weakness." I concur with this observation. One of the unrecognized advantages of formal models like Powell's is that they facilitate the exposure of underlying assumptions to careful scrutiny.

Beyond this, however, there is yet another problem with attributing the stability of the superpower relationship during the postwar period to each side's use of the strategy-that-leaves-something-to-chance: neither superpower has ever acted this way, at least in the nuclear age. As Young finds in an empirical examination of four post-World War II crises, in such situations each side acted "to retain wide freedom of choice as long as possible and to avoid becoming boxed into an irrevocable position."[55] Thus, not only have the superpowers eschewed commitment strategies like those suggested by Gauthier or Brams and Kilgour, but they have also avoided, in actual crisis situations, those brinkmanship tactics modeled by Powell which, by forfeiting control, unleash an autonomous risk of war.

Toward a Resolution

Can deterrence theory be rescued from this theoretical trap by explaining deterrence stability and simultaneously maintaining consistency with the instrumentalist definition of rationality? A definitive answer is beyond the scope of this essay. Nevertheless, a sketch of a potential resolution is feasible.

One of Ellsberg's suggestions provides a useful starting point: if a deterring player can convince his opponent that the costs of capitulation are actually higher than the costs of executing the deterrent threat, then execution of the threat will, in fact, be rational.[56] That "second wave" deterrence theorists have found this suggestion compelling is revealed by the numerous strategems they have devised for doing just this. These tactics range from making a public commitment to retaliation, thereby raising the costs of capitulation, to linking the issue at hand to future confrontations or to such intangibles as "national honor." The most ingenious tactic, however, is Schelling's notorious suggestion that a player feign irrationality by appearing to be oblivious to the costs of confrontation. Notice, however, that the "rationality of irrationality" strategy turns upon a denial of procedural (though significantly not instrumental) rationality. The player feigning irrationality will appear to be procedurally irrational because his preferences will appear to be different than those deemed reasonable by deterrence theorists; but this player will be instrumentally rational since he is presumed to act consistently with these irrational preferences.

I have shown elsewhere that if each player is able to convince his op-

[55] Oran R. Young, *The Politics of Force: Bargaining During International Crises* (Princeton, NJ: Princeton University Press, 1968), 218.

[56] Ellsberg (fn. 13), 357.

ponent that he actually prefers confrontation to capitulation, thereby transforming the underlying Chicken game into a game with the structural characteristics of "Prisoners' Dilemma" (see Figure 4), then a tenuous deterrence equilibrium can emerge, provided that strategy choices in this game are made both sequentially and conditionally, and that each player has an invulnerable second-strike capability.[57] Notice that this resolution of the dilemma satisfies not only the perfectness criterion but also is consistent with the prescriptions of most deterrence theorists who argue that deterrence works best when each side possesses a credible (i.e., rational) retaliatory threat.

		Nation B	
		Cooperate (C)	Not Cooperate (D)
Nation A	Cooperate (C)	(3,3)	(1,4)
	Not Cooperate (D)	(4,1)	(2,2)

Figure 4
Prisoners' Dilemma

The problem here, however, is that the resolution raises fundamental questions about the underlying determinants of preferences. As Ellsberg himself put it, "Whose reputation for honesty is so great that to wager it would make it actually *rational* to carry out such a threat? And who, with such issues at stake, would really ... carry out a suicidal punishment?"[58] Or as de Gaulle once asked, rhetorically, "Would the United States trade New York for Paris?" The answer is that no one, including analysts or statesmen, knows. But the reason we do not know is not because we do not know whether the players will be instrumentally rational. Both proceduralists and instrumentalists assume that they will be. We do not know because we are ignorant about the conditions under which statesmen are prepared to fight or capitulate, that is, we do not as yet have an understanding of the forces which give rise to the preferences of the players. If we did, the resulting game (of complete information) would be fully specified and, as Wagner has persuasively demonstrated,

[57] Zagare (fn. 2, 1987).
[58] Ellsberg (fn. 13), 358.

completely determined.[59] This is precisely why Schelling and most students of interstate conflict hold that the essence of a crisis is its unpredictability. When preferences are given, deterrence either works or it does not.[60] Powell's model supports this: when information is complete, there is no crisis equilibrium.

Since deterrence is undoubtedly a game of incomplete information, the critical question becomes whether or not a deterrence equilibrium can emerge under these conditions. Elsewhere, Kilgour and I have examined just such a game and—not surprisingly—have discovered that it can.[61] In formulating this model, we assume that each player is uncertain about his opponent's preferences should he unilaterally attempt to alter the status quo. In turn, uncertainty about the preferences of one's opponent leads to uncertainty about the opponent's willingness to retaliate. Since the credibility of each player's threat is identified with the probability that a player prefers retaliation to capitulation, this model maintains consistency with both the traditional strategic literature, in which credibility is usually equated with believability, and with the literature of game theory, in which credibility is usually taken to be synonymous with sequentially rational (i.e., subgame perfect equilibrium) choices.

This is clearly not the proper forum for giving the details of the abovementioned deductions. Suffice it to say that the model reveals that when the credibility of each player's threat is "sufficiently" high, deterrence is likely, though not necessarily certain, as some theorists have speculated.[62] Moreover, the existence of a deterrence equilibrium depends on each player's evaluation of the status quo and the costs of conflict. Interestingly, however, no linear relationship exists between the absolute costs of warfare and deterrence stability. In fact, the model demonstrates that in core areas where both players have inherently credible threats, increasing the costs of conflict past a certain point does little to enhance deterrence stability.

Even if one is persuaded by the synopsis of this resolution to the paradox of mutual deterrence—and one need not be—it is not the end of the story. The research question that remains unanswered, even at this

[59] Wagner (fn. 2, 1982), 343.

[60] Specifically, if each player prefers executing the deterrent threat, nuclear deterrence is stable; if neither player has a credible threat, it is not; and if only one player's retaliatory threat is credible, that player wins (Zagare [fn. 2, 1987], chap. 2).

[61] D. Marc Kilgour and Frank C. Zagare, "Uncertainty and Deterrence" (Paper delivered at the Annual Meeting of the American Political Science Association, Atlanta, GA, August 31-September 3, 1989).

[62] Michael D. Intriligator and Dagobert L. Brito, "Nuclear Proliferation and the Probability of Nuclear War," *Public Choice* 37, No. 2 (1984), 247-60; John Mueller, *Retreat From Doomsday: The Obsolescence of Major War* (New York: Basic Books, 1989).

point, concerns the basis and origins of the perceptions, correct or not, of an opponent's willingness to execute the deterrent threat. Obviously, the answer to this question cannot be determined wholly within the confines of the rational choice paradigm but rather falls also within the competency of those theorists whose work lies in the tradition of procedural rationality. Thus we come full circle. The argument here is that deterrence theory is perfectly consistent with both the ostensively antirationalist tradition of cognitive research based on the notion of procedural rationality and the tradition of the instrumental strategist. Perhaps a more meaningful understanding of international conflict can be achieved if these two schools realize that they are involved in a nonzero-sum research quest which, at present, is characterized by cognitive closure on the part of the proceduralists and procedural irrationality on the part of rational choice deterrence theorists.