

RATIONAL DETERRENCE THEORY: I Think, Therefore I Deter

By RICHARD NED LEBOW and JANICE GROSS STEIN*

... abstract models are like scaffolding used to build a structure; the structure must stand by itself. If the abstract models contain empirical falsities, we must jettison the models, not gloss over their inadequacies.

—Paul Samuelson¹

PROponents of “rational deterrence theory” contend that it is conceptually sound, a good predictor of strategic behavior, and a successful strategy of conflict management. All three assertions are unwarranted. Existing theories of deterrence are incomplete and flawed. They are poor predictors of critical cases of strategic behavior and equally poor guides to policy. Theories of deterrence do not consider the most important determinants of strategic choice. These determinants are outside of and at times contradictory to their fundamental assumptions. To develop theories with predictive capability and policy relevance, scholars must go beyond deterrence and rational choice to other theories of international behavior.

THEORIES OF RATIONAL CHOICE

Theories of rational choice assume that people maximize expected utility. But rational-choice theorists recognize that utility can be calculated differently by different but equally rational actors.² Some people may rationally minimize expected losses while others may maximize expected gains. People can also be risk-prone or risk-averse without deviating

*The research and writing of this article were supported by grants from the John T. and Catherine D. MacArthur Foundation and the Canadian Institute for International Peace and Security.

¹ Samuelson, “Problems of Methodology—Discussion,” *American Economic Review* 53 (May 1963), 236.

² This problem is discussed by Frank C. Zagare in “Recent Advances in Game Theory and Political Science,” in Samuel Long, ed., *Annual Review of Political Science* (Norwood, NJ: Ablex Publishing Corporation, 1986), 60-90, and in *The Dynamics of Deterrence* (Chicago: University of Chicago Press, 1987), 1-7. It is also a central theme of Arthur Stein, *Dilemmas of Interdependence: The Logic of International Cooperation and Conflict*, forthcoming.

from the norm of rationality.³ Subjective expected utility will vary enormously depending on actors' risk propensity and relative emphasis on loss or gain.⁴ Risk-prone, gain-maximizing leaders—Hitler comes readily to mind—make very different choices than their risk-averse, loss-minimizing counterparts.

Models of rational choice must therefore build in assumptions about the risk propensities of players and the relative weighting of gains and losses. To do so they must turn to other theories of behavior.⁵ As Herbert Simon has pointed out, a rational political theory is always more than rational utility theory. The principle of rationality says very little that is interesting or important.⁶ It only takes on meaning through auxiliary propositions that can provide the basis for a deductive theory of political behavior.⁷

THEORIES OF DETERRENCE

Some theories of deterrence purport to supply the additional assumptions rational-choice theories need to predict rational strategic behavior. Theories of deterrence generally assume that would-be initiators are risk-prone gain maximizers. Christopher Achen and Duncan Snidal properly describe the principal explanatory assumption of deterrence theories as the expectation that variation in outcomes can be explained by differences in actors' opportunities.⁸ The fundamental theoretical concern of deter-

³ Formal concepts of risk and uncertainty apply to the estimation of the likelihoods of given alternatives. In a "risky" choice, the probability distribution of a consequence is known, even though its occurrence is not. In an uncertain environment, by contrast, even this distribution of probabilities is unknown and no a priori estimates can be made. The formal distinction between risk and uncertainty does not capture the attributes of the environment of strategic decision making. Once leaders estimate the probability of a consequence of an option, even without adequate baseline information on which to make their judgments, they make their choice under conditions of risk.

⁴ In models of subjective expected utility, preferences under conditions of risk are determined by at least two separate factors: the strength of the preference for the consequences under certainty, and the attitude toward risk. See Paul J. H. Schoemaker, "The Expected Utility Model: Its Variants, Purposes, Evidence, and Limitations," *Journal of Economic Literature* 20 (June 1982), 529-63, at 533.

⁵ Joseph S. Nye, Jr., develops this argument in "Neorealism and Neoliberalism," *World Politics* 40 (January 1988), 235-51.

⁶ Simon, "Human Nature in Politics: The Dialogue of Psychology with Political Science," *American Political Science Review* 79 (June 1985), 293-304.

⁷ For a cogent exposition of this argument, see Dwain Mefford, "An Alternative to Rational Choice: The Function of Stories in the Formulation of Foreign Policy with Examples from U.S. Intervention," paper delivered at the MIT-Harvard seminar, "Rethinking Security Relationships and International Institutions and Cooperation," Cambridge, MA, May 5, 1988.

⁸ Achen and Snidal, "Rational Deterrence Theory and Comparative Case Studies," *World Politics* 41 (January 1989), 150.

rence has therefore been a specification of the ways in which defenders can use threats to increase leaders' calculations of the likely costs of a challenge.

This assumption of deterrence is arbitrary. It ignores the three other logical possibilities: initiators could also be risk-prone loss minimizers, risk-averse gain maximizers, and risk-averse loss minimizers. They are not all like Adolf Hitler. Nikita Khrushchev was risk-prone but appears to have emphasized loss, as were Japan's leaders in 1941, and Anwar el-Sadat in 1973.⁹ Joseph Stalin is probably best described as a risk-averse gain maximizer.¹⁰

Empirical analyses of deterrence failures indicate that risk-prone gain-maximizing initiators—the assumption of almost all theories of deterrence—are relatively uncommon. The most serious challenges to deterrence most often come from initiators who are risk-prone but emphasize loss. Case studies indicate that outside threats have no effect on the calculations of this kind of initiator or may make a challenge even *more* attractive.¹¹

⁹ See Richard Ned Lebow, "Provocative Deterrence: A New Look at the Cuban Missile Crisis," *Arms Control Today* 18 (July-August 1988), 15-16, for recent Soviet revelations about Khrushchev's motives for the Cuban missile deployment. Earlier Western studies that emphasized Khrushchev's concern for avoiding loss include, Arnold Horelick and Myron Rush, *Strategic Power and Soviet Foreign Policy* (Chicago: University of Chicago Press, 1966), 141; Roger Hilsman, *To Move A Nation: The Politics of Foreign Policy in the Administration of John F. Kennedy* (Garden City, NY: Doubleday, 1967), 200-202; and Jerome H. Kahan and Anne K. Long, "The Cuban Missile Crisis: A Study of Its Strategic Context," *Political Science Quarterly* 87 (December 1972), 564-90. On Japan, see Robert Butow, *Tojo and the Coming of the War* (Stanford: Stanford University Press, 1961); N. Ike, *Japan's Decision for War, Records of 1941: Policy Conferences* (Stanford: Stanford University Press, 1967); C. Hosoya, "Miscalculations in Deterrence Policy: Japanese-U.S. Relations, 1938-1941," *Journal of Peace Research* 5, No. 2 (1968), 97-115; Bruce Russett, "Deterrence Theory and Decision Theory," *Journal of Peace Research* 4, No. 2 (1967), 89-105; D. Borg and S. Okamoto, eds., *Pearl Harbor as History: Japanese-American Relations, 1931-1941* (New York: Columbia University Press, 1973); and S. Ienaga, *The Pacific War, 1931-1945* (New York: Pantheon, 1978). On Sadat, see Janice Gross Stein, "Calculation, Miscalculation, and Conventional Deterrence I: The View from Cairo," in Robert Jervis, Richard Ned Lebow, and Janice Gross Stein, *Psychology and Deterrence* (Baltimore: The Johns Hopkins University Press, 1985), 34-59.

¹⁰ See Vojtech Mastny, *Russia's Road to the Cold War: Diplomacy, Warfare, and the Politics of Communism, 1941-45* (New York: Columbia University Press, 1979), and William Taubman, *Stalin's America Policy: From Entente to Detente to Cold War* (New York: Norton, 1982), for this interpretation.

¹¹ For an analysis of the conditions under which the strategy of deterrence is likely to provoke rather than prevent a use of force, see Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976), 58-113, and "Rational Deterrence: Theory and Evidence," *World Politics* 41 (January 1989), 183-207; Richard Ned Lebow, *Between Peace and War: The Nature of International Crisis* (Baltimore: The Johns Hopkins University Press, 1981), and "The Cuban Missile Crisis: Reading the Lessons Correctly," *Political Science Quarterly* 98 (Fall 1983), 431-58; Richard Ned Lebow and Janice Gross Stein, "Beyond Deterrence," *Journal of Social Issues* 43, No. 4 (1987), 3-71; and Janice Gross Stein, "Deterrence and Miscalculated Escalation: The Outbreak of War in 1967," paper presented to the American Political Science Association, September 1988.

Deductive theories of deterrence that assume that challengers are risk-prone gain maximizers will not only be poor predictors of outcomes but will also create serious problems for those who use the theory as a guide to strategy. Recognizing this problem, some analysts have begun to experiment with varying assumptions of risk propensity in their models of subjective expected utility.¹² If these efforts are to be grounded in theory, analysts must look beyond deterrence to other theories of international behavior for guidance. Ultimately, however, the risk propensity of challengers and their relative weighting of loss and gain are empirical questions that can be answered only by careful historical research.

Deductive theories of deterrence also err in their specification of the critical factors that influence a challenger's calculus. Achen and Snidal, once again representative of deterrence theorists, argue that the calculus of deterrence is straightforward and undemanding: "If a country knows that it is likely to lose a long, nasty war in the process, it will probably not seek to press its claims against a rival. The trick is to learn the likelihood that the rival country will fight—and if it fights, how likely it is to win." More formally, they suggest that an initiator's choices depend upon its subjective estimates of the costs of war, the probability of winning, and the estimated probability that the defender will retaliate.¹³

Empirical analyses of deterrence failures have identified cases in which leaders calculated in accordance with the expectations of deterrence theories but acted contrary to their predictions. These leaders estimated the expected costs of war as very high, the probability of winning as low, and the probability that the defender would retaliate as virtually certain. Yet, they chose to challenge deterrence. Japan's leaders made these kinds of calculations when they chose to attack the United States in 1941, as did Sadat when he decided to challenge Israel in 1973. Contrary to the assertion of Achen and Snidal, cases of this kind do constitute a partial test of the predictive validity of "rational deterrence theory." They demonstrate that the interrelationships among the variables specified by the theory did not produce the anticipated outcome.

A close examination of these and other cases of deterrence failure indicates that the calculus of initiators depends on factors other than those identified by deterrence.¹⁴ Perhaps the most important of these is the expected domestic political consequences of a use of force. For Sadat in

¹² Bruce Bueno de Mesquita and David Lalman, "Reason and War," *American Political Science Review* 80 (December 1986), 1113-30; Stein (fn. 2).

¹³ Achen and Snidal (fn. 8), 151 and n. 24.

¹⁴ These cases are analyzed in Jervis, Lebow, and Stein (fn. 9), 34-59, 203-32, and in Lebow and Stein (fn. 11).

1973, the critical consideration—and a variable omitted from deterrence theories—was his estimate of the adverse economic and political consequences of inaction.

Rational deterrence theories do more than omit important variables. Theories of deterrence focus on the manipulation of an initiator's estimates of the costs of war, the probability of winning, and the probability of retaliation through credible threats reinforced by usable military options. Once analysts admit the critical relevance of other variables, it becomes necessary to determine the relative weighting of these variables or the circumstances in which one or another dominates calculations of utility. Achen and Snidal admit the importance of domestic politics but neither they nor theories of deterrence provide a basis for the specification and weighting of this and other omitted variables. The relative importance of the domestic consequences of action or inaction cannot be derived from any of the assumptions, postulates, or axioms of "rational deterrence theory," no matter how they are reformulated. We must once again turn to other theories of international behavior.

CAN "RATIONAL DETERRENCE THEORY" BE SAVED?

How do proponents of "rational deterrence theory" accommodate evidence of its failure to predict outcomes in important cases? They acknowledge its weaknesses, relax critical assumptions, and fall back on a broad definition of net subjective expected utility as the predictor of deterrence outcomes. Achen and Snidal, like many others, assert that the initiator's subjective expected utility of attacking must be less than that of continuing with the status quo.¹⁵ This is an unworkable theoretical formulation.

Whether or not an actor is rational is beside the point. Deterrence theory does not predict that initiators will be rational. It specifies the conditions under which rational initiators will choose not to attack. The two are not the same. Deterrence theory derives its rigor and predictive power from its specification of the interrelationship among the critical variables that allegedly shape an initiator's calculus. If its claim to theory is to be taken seriously, the variables it identifies and their interaction must have the anticipated outcome. When a challenger makes a decision to use or not to use force with reference to criteria outside of deterrence theory—such as domestic or alliance politics—then the validity of deterrence theory is doubtful.

¹⁵ Achen and Snidal (fn. 8), 152, n. 24; Philip E. Tetlock, "Testing Deterrence Theory: Some Conceptual and Methodological Issues," *Journal of Social Issues* 43, No. 4 (1987), 85-92.

More damning still, deductive theories of deterrence, even those built on revised models of net subjective expected utility, do not define their scope conditions. Achen and Snidal extol the virtues of a simple formulation of subjective expected utility: if the expected punishment exceeds the gain, they argue, then opponents will be deterred. They concede that deterrence will fail against sufficiently determined attackers.¹⁶ Recognition that some initiators are not deterrable, while obvious, does little to rescue deterrence theories. A theory worth the name must specify the conditions of its relevance. If it does not, it cannot be tested against any body of evidence, no matter how complete.¹⁷ It also runs the risk of becoming self-confirming because anomalous results can be interpreted after the fact as outside the scope of the theory.¹⁸

Achen and Snidal's definition of the scope conditions of "rational deterrence theory" is even more confusing than the deceptively simple formulation of the net subjective expected utility of attack suggests. They assert that deterrence theory predicts that deterrence will fail when a deterrer's threat of denial or punishment is absent, incredible, or less valuable than the prize. But it should be obvious that when no threat is made, deterrence is not relevant; these cases lie outside the scope of the theory. The theory is also mute about the conditions in which initiators are likely to judge the prize more valuable than the expected costs of a challenge. It speaks only, and then only in part, to the conditions that make a threat credible.

While the consequences of the failure to formulate scope conditions are serious for the theory of deterrence, their implications are even more troubling for the strategy of deterrence. Leaders can get no guidance on when an attempt at deterrence is appropriate, when it is likely to fail, and when it is likely to provoke an incautious and vulnerable adversary. Be-

¹⁶ Achen and Snidal (fn. 8), 152.

¹⁷ In reviewing the predictive use of subjective expected utility models, Schoemaker (fn. 4, 538-39) argues that evidence counter to the theory can often be refuted because it is indirect. The most irrefutable evidence, which directly concerns the axioms, is discounted in the positivist tradition. Ex post facto interpretation of a decision as rational through the addition of factors not specified by the model treats the optimality of behavior as an essentially unfalsifiable metapostulate. Consequently, ex post empirical models have limited refutation power regarding the corresponding theoretical *ex ante* model.

¹⁸ Mefford (fn. 7) argues persuasively: "Interpreting the *uninterpreted* key concepts of rational choice requires spelling out for each actor that actor's utility function, perceptions of uncertainty, and perceived set of outcomes. Because there is virtually no limit on the variety of ways that these concepts of rational choice can be mapped onto a political world, rational-choice based models of deterrence seem to suffer from much the same open-ended character . . . of the case-based work. . . . The number of auxiliary propositions that can be attached to the uninterpreted core of rational choice is disturbingly large. It matters which ones are chosen. And, the rational choice theorist's choice of auxiliary propositions is a legitimate target for critical examination" (emphasis in original), p. 9.

cause deductive theories of deterrence do not specify their scope conditions, leaders are encouraged to inflate the strategy of deterrence to a general strategy of foreign policy.¹⁹

A better specified theory that solved these problems would no longer be a theory of deterrence. It would have to be based on a much broader set of assumptions about an initiator's calculus and the additional factors likely to influence that calculus. It would incorporate a wider range of strategies appropriate to these assumptions. In some circumstances, strategies of reassurance, which attempt to reduce the incentives initiators have to use force, may be far more effective than deterrence in avoiding war.²⁰ Insofar as we go outside theories of deterrence, we may build a better theory of strategic choice, but *not* of deterrence, which is limited to the manipulation of credible threat.

THE LIMITS OF RATIONAL CHOICE

Theories of deterrence built on better models of subjective expected utility or game-theoretic analyses of strategic preferences would still suffer from crippling limitations.

Neither theories of deterrence nor rational choice say anything about the all-important preferences that shape leaders' calculations.²¹ Achen and Snidal correctly observe that deterrence theory assumes exogenously given preferences and choice options. It begs the question of how preferences are formed. Empirical analyses of decision making suggest that individuals often identify their preferences and options in the course of formulating and reformulating a problem. Herbert Simon has argued:

To understand political choices, we need to understand where the frame of reference for the actors' thinking comes from—how it is evoked. An important component of the frame of reference is the set of alternatives that are given consideration in the choice process. We need to understand not only how people reason about alternatives, but where the alternatives come from in the first place.²²

The most important determinant of strategic decisions is not the process of choosing among options but the prior definition and construction of the problem to be decided. Although it is legitimate for theory to set

¹⁹ For a similar argument without reference to the analytical properties of the theory, see Michael McGwire, "Deterrence: The Problem—Not the Solution," *SAIS Review* 5 (Summer-Fall 1985), 105-24.

²⁰ For analysis of the interaction between strategies of deterrence and reassurance in avoiding war, see Janice Gross Stein, "Deterrence and Reassurance," in Philip E. Tetlock et al., eds., *Behavior, Society, and Nuclear War* (New York: Oxford University Press, forthcoming).

²¹ Jervis (fn. 11, 1989).

²² Simon (fn. 6), 295.

its own boundaries, the boundaries set by “rational deterrence theory” exclude most of the theoretically important problems of deterrence and the politically relevant questions leaders confront when they use it as strategy.²³

Of equal importance, “rational deterrence theory” acknowledges that preferences or estimates change but says nothing about how, when, and why this occurs. The assumption that initiators change their preferences and estimates as a consequence of the actions of the defender lies at the very core of theories of deterrence. Much of the game-theoretic analysis of deterrence posits an ordering of preferences and predicts behavior accordingly. Yet preferences often change through the process of interaction and may do so in ways that are unanticipated by or contradictory to theories of deterrence.

The Cuban missile crisis offers a striking example of this phenomenon. The use of threats by President Kennedy to moderate Soviet behavior unexpectedly increased Soviet estimates of the strategic costs of inaction and shifted the preference ordering of Soviet leaders so as to make a challenge more attractive. Soviet sources report that Kennedy’s strategic buildup, his decision to proceed with the deployment of Jupiter missiles in Turkey, and his claims of strategic superiority convinced Khrushchev and Marshal Rodion Malinovsky that it was imperative to do something to improve the Soviet strategic position.²⁴ Not only did learning occur, but the learning that did occur was contradictory to the predictions of deterrence theories. For both these reasons, the analysis of deterrence in theory and practice is hollow at its center.

A different and more difficult problem for theories of rational choice is the now-well-documented deviations from processes of rational calculation. At times, these can be so severe and pervasive as to defeat deterrence in ways unanticipated by the theory. There is strong evidence from comparative analysis of cases of deterrence failure that initiators frequently distort information about the expected costs of military action, the probability of winning, and the probability that the defender will retaliate—the three variables identified as critical by “rational deterrence theory.”²⁵ Leaders in fact deviate systematically from rational norms in their estimates of the factors specified by “rational deterrence theory” as critical to leaders’ calculations.²⁶ Misperception and miscalculation are

²³ For a similar argument with respect to game theory and the analysis of cooperation, see Robert Jervis, “Realism, Game Theory, and Cooperation,” *World Politics* 40 (April 1988), 317-49.

²⁴ Lebow (fn. 9), *passim*.

²⁵ Lebow and Stein (fn. 11), *passim*.

²⁶ As Mefford argues (fn. 7), what is at issue is the correspondence between the essential terms in the core of rational models and the auxiliary assumptions of deterrence theory (pp.

important not only as descriptions of how leaders think, but also as predictors of what leaders choose.

Two kinds of evidence make clear the consequences of omitting the impact of misperception and miscalculation from the analysis of deterrence. The first is drawn from the analysis of a critical case, the second from the results of a deductive analysis premised on "rational deterrence theory." In May 1967, Egypt challenged deterrence because its leaders seriously underestimated the costs of military action and overestimated the probability of winning. They did so even though only a few weeks earlier they had been pessimistic about the probability of winning and skeptical of their military capabilities. The dramatic change in their estimates of relative capabilities cannot be explained by objective changes in the balance of power but rather by motivated distortion in response to pressing political needs.²⁷

The "rational theory of deterrence," applied to Egyptian preferences at the beginning of May, would have predicted the success of deterrence. Alternatively, applied to the miscalculated estimates of Egyptian leaders in mid-May, it would have predicted the failure of deterrence, but for reasons the theory does not consider. Egyptian estimates did not change in response to the behavior of the deterrer, but largely in reaction to pressing needs at home and in the Arab world. Theories of deterrence would not have been able to capture or explain, much less predict, the changes in the estimates that determined Egyptian choices. Almost everything relevant, for strategy as well as for theory, is outside their scope.

This fundamental inadequacy of deterrence is acknowledged by Frank Zagare, a leading proponent of game-theoretic analyses.²⁸ To explain the initial success of the United States in deterring Israel from using military force on May 28, 1967, he explicitly uses Israel's miscalculated estimates of American preferences. In his modeling of the failure of American deterrence only one week later, he works with Israel's accurate estimates of American preferences. Zagare's theory of moves first builds in miscalculation as an exogenous variable to explain the success of deterrence. Its analysis of deterrence failure a week later cannot account for the revision of the estimates that changed the outcome.²⁹ Again, what is

7-10). These auxiliary assumptions are the costs of war, the probability of winning, and the probability of retaliation. Achen and Snidal ignore the necessity to test these propositions as the core of an empirical theory of deterrence.

²⁷ Stein (fn. 11).

²⁸ Zagare (fn. 2, 1987), 1-7.

²⁹ Technically, Zagare's analysis is based on a different interpretation of the meaning of two-by-two games. The numbers denote rates of payoff rather than single payoffs, with players moving around within the matrix.

critically relevant to theory and strategy is outside the model. A recent review of game-theoretic analyses of deterrence acknowledges this critical weakness and urges that rational actions be treated in part as a function of rational beliefs.³⁰

“Rational deterrence theory” cannot by definition accommodate systematic deviations from rational choice. Deterrence theorists are trying to cope with this dilemma by utilizing recent advances in game theory that incorporate uncertainty and sequential equilibria into their models.³¹ But these models do not address the problems of misperception and miscalculation that we have identified. Uncertainty is the result of inadequate information. Misperception and miscalculation arise from faulty evaluation of available information. Studies of deterrence and intelligence failures find that error rarely results from inadequate information but is almost always due to theory and stress driven interpretations of evidence.³² On this one, “rational deterrence theorists” are barking up the wrong decision tree.

Achen and Snidal assert that psychopathological reasoning lies outside the scope of rational deterrence theory. But they and others fail to define the boundaries of such reasoning.³³ More generally, deterrence theorists have failed to delineate the boundaries of rational choice: if the problem of deterrence theory is defined as whether or not a rational actor will choose to attack despite a credible threat of retaliation, then specifying the

³⁰ Barry O'Neill, “Game Theory and the Study of the Deterrence of War,” in Paul C. Stern et al., eds., *Perspectives on Deterrence* (New York: Oxford University Press, 1989). O'Neill notes: “Previous concepts had tried to draw a sharp line between rational and irrational actions based solely on the utilities of the outcomes, but recent approaches regard rational actions as interlocked with rational beliefs. . . . Whether they [the players] adopt these beliefs is an empirical question, and thus the validity of an equilibrium becomes a question of social science rather than mathematical axioms. Models like this seem more realistic and reduce the isolation of game theory applications from other social sciences.”

³¹ See, for example, Robert Powell, “Crisis Bargaining, Escalation, and MAD,” *American Political Science Review* 81 (September 1987), 717-36, and “Nuclear Brinkmanship with Two-Sided Incomplete Information,” *American Political Science Review* 82 (March 1988), 155-78; R. Harrison Wagner, “Reputation and the Credibility of Military Threats: Rational Choice vs. Psychology,” paper presented to the Annual Meeting of the American Political Science Association, September 1988.

³² Harvey de Weerd, “Strategic Surprise and the Korean War,” *Orbis* 6 (Fall 1962), 435-52; Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974), 184-234; Lebow (fn. 11, 1981) passim; Janice Gross Stein, “Calculation, Miscalculation, and Conventional Deterrence II: The View from Jerusalem,” and Richard Ned Lebow, “Miscalculation in the South Atlantic: The Origins of the Falkland War,” in Jervis, Lebow, and Stein (fn. 9), 60-88, 89-124.

³³ The Egyptian Minister of War, Shams al-Din Badran, estimated in May 1967 that Egyptian forces could handle military intervention on behalf of Israel by the Sixth Fleet; cited by Mahmoud Riad, *The Struggle for Peace in the Middle East* (London: Quartet Books, 1981), 23. Is this, in Snidal and Achen's terms, psychopathological reasoning that falls outside the scope of a “rational theory of deterrence,” or evidence of systematic miscalculation that defeats the expectations of rational deterrence theory?

criteria of rationality is absolutely critical to determining the scope of the theory.

Are deductive theories of deterrence restricted to cases where leaders rationally calculate subjective expected utility or reasonably approximate these norms? How much do leaders have to deviate from norms of rationality to be considered nonrational? When initiators systematically discount the likely costs of military action because the choices they confront are so painful, or dismiss clear evidence of the likelihood of retaliation by a defender, or ignore the central trade-offs among alternatives, are they acting outside the scope of "rational deterrence theory"? If they are, then "rational deterrence theory" is restricted to a small proportion of the cases where deterrence is at issue. A general review of subjective expected utility theory concludes that it fails as a predictive and descriptive model in almost all the cases where it has been empirically tested. This is due to its inadequate specification of psychological principles of judgment and choice.³⁴ Rational deterrence theory may be elegant but irrelevant.

This problem is not unique to deterrence theory but applies to rational-choice theories generally. Rational-choice theorists confront an awkward dilemma. They can pursue elegance at the price of relevance, or they can reject flawed assumptions and reformulate their theories. Paul Schoemaker complains that they have chosen the former by trying to save their theories by tinkering with one or two peripheral assumptions. "Having invested heavily in complex deductive structures, with wide domains of applicability and mathematically elegant decision models that allow for easy aggregation across people, it is a natural tendency to patch up the theory cosmetically. A better alternative," he insists and we agree, would be "to examine closely the type of anomalies reported and the cognitive reasons underlying them."³⁵

TESTING THEORIES OF DETERRENCE

Achen and Snidal do not dispute the evidence of systematic deviations from rational choice identified in the comparative analysis of cases of deterrence failure. They defend "rational deterrence theory" with the assertion that its purpose is not to model leaders' calculations but to predict outcomes. They insist that if deterrence theory makes accurate predictions, evidence of constraints to rationality is of no concern.

Leaving aside for the moment the claim that good theory can be built

³⁴ Schoemaker (fn. 4), 548.

³⁵ *Ibid.*, 554.

on flawed behavioral assumptions, there is no justification for the assertion that deterrence theory has predictive value. The predictive capacity of rational deterrence theory has never been demonstrated nor satisfactorily tested. What evidence there is, indicates that it is a poor predictor of behavior in critical cases.

How has deterrence theory been tested empirically? Some scholars have used aggregate data analysis to test the predictive power of the theory in cases of general deterrence.³⁶ In a study of extended general deterrence cited approvingly by Achen and Snidal, Erich Weede identifies all logically possible cases of deterrence, without examining whether the criteria of deterrence are present, and then finds significant differences in the incidence of war.³⁷ The study fails to define the scope conditions of the theory and does not investigate empirically whether or not deterrence was relevant to the thousands of cases it identifies. By the nature of his case selection, many if not most of his cases, we suspect, fall outside the scope of deterrence. If in these cases neither party contemplated a challenge nor sought to deter one, the study dramatically overestimates the success of deterrence.

Paul Huth and Bruce Russett are more rigorous in defining what constitutes an immediate deterrence encounter and have identified sixty cases of extended immediate deterrence in this century.³⁸ Under the most permissive criterion of success, deterrence succeeds only 59 percent of the time.

To determine the causes of success and failure, Huth and Russett use indirect measures of alliance commitments, balance-of-power ratios, and past bargaining behavior to tap indirectly leaders' estimates of commitment, capability, and credibility. They then use probit analysis to identify the structural relationships between these variables and the outcome of deterrence. Their method allows them to estimate the relationship between structural and behavioral conditions and the outcome of deterrence but not to test directly the propositions deduced from deterrence theory. They nevertheless acknowledge the importance of directly tracing the impact of leaders' estimates on the outcome of deterrence.³⁹

³⁶ For the distinction between "immediate" and "general" deterrence, see Patrick Morgan, *Deterrence: A Conceptual Analysis*, rev. ed. (Beverly Hills: Sage Publications, 1983).

³⁷ Weede, "Extended Deterrence by Superpower Alliance," *Journal of Conflict Resolution* 27 (June 1983), 231-54.

³⁸ Huth and Russett, "What Makes Deterrence Work? Cases from 1900 to 1980," *World Politics* 36 (July 1984), 496-526, and "Deterrence Failure and Crisis Escalation," *International Studies Quarterly* 32, No. 1 (1988), 29-46; Paul Huth, "Extended Deterrence and the Outbreak of War," *American Political Science Review* 82 (June 1988), 423-44, and *Extended Deterrence and the Prevention of War* (New Haven: Yale University Press, 1988).

³⁹ Russett (fn. 9), and Huth (fn. 38).

Our empirical research has concentrated on immediate deterrence failures. We chose to work with failures because they are more easily identifiable through the observable behavior that occurs. Achen and Snidal question whether this approach can test deterrence theory. They claim that it can only identify and describe failures of deterrence as a strategy. They are wrong. We define the scope conditions of our cases to conform to the expectations of theories of deterrence and are therefore able to identify failures of theory as well as of strategy.

We distinguish between two kinds of deterrence failure: those that present no problem to deterrence theory and those that do. The former consists of cases where deterrence failed but where the strategy was implemented poorly or improperly. We draw no inferences about deterrence theory from these cases. The latter set of cases have damaging implications for the theory. In all of them, the defender met the conditions of deterrence theory: the deterrer carefully defined the unacceptable behavior, threatened retaliation, had the capability to implement the threat, demonstrated resolve, but the challenger still proceeded to use force. In some of these cases, the contradiction between theory and behavior can be attributed to the flawed calculations of the challenger. Our empirical evidence indicates that this is a widespread phenomenon. A second kind of failure occurs when challengers calculate as deterrence theory expects but behave contrary to its predictions. They do so because their calculations are determined by considerations outside the realm of deterrence theory. Cases of this kind stand as a stark challenge to that theory.

Ours is a highly restrictive definition of failure, for it deliberately excludes cases where challengers resorted to force but defenders did not attempt to deter or did so ineptly. Unlike Achen and Snidal, we do not consider that deterrence has failed when the threat of denial or punishment is absent.⁴⁰ Because we base our analysis only on cases that meet the conditions of deterrence theory, we can claim to test for failures of that theory.

We do not contend that our analysis of immediate deterrence failures encompasses the universe of cases, either of deterrence failures or successes. This is in large part because of the data requirements of the analysis of deterrence. To identify cases, we need valid evidence of the calculations of both the initiator and the deterrer. We seek, whenever possible, not only the reconstructions of the participants after the fact, for, as Achen and Snidal note and we concur, this kind of evidence is subject to well-known biases. We look for convergent evidence from several par-

⁴⁰ Achen and Snidal (fn. 8), 152.

ticipants from each side, and for historical documentation as well. We remain unpersuaded that indirect measures used in the aggregate analysis of deterrence are valid measures of the variables specified by theories of deterrence.⁴¹ These criteria of reliability and validity of evidence sharply limit the number of cases suitable for analysis. Fortunately, for many of the critical cases of deterrence failure, better evidence on the calculations of both sides is becoming available.

Detailed analysis of cases reveals a problem that is ignored by deterrence theory. This is the difficulty of distinguishing challenger from defender in any given case.⁴² Deterrence theories assume that there is a clear analytical distinction between challenger and defender that is obvious not only to scholars but to the actors themselves. In reality, this distinction is generally blurred. Both parties to a deterrence encounter often consider themselves to be the defender. This is not just a technical problem, but one with significant theoretical consequences. Leaders' conceptions of themselves as initiators or defenders have important consequences for their behavior. It also makes coding of deterrence cases very difficult and at times arbitrary. The 1914 crisis is an obvious case in point. So too are the Cuban missile crisis and Egypt and Israel in 1967.

More important, even if we were able to identify the universe of cases and determine who was the challenger and who the defender, establishing the proportional rate of deterrence success and failure is not the only important question. Even a single deterrence failure can lead to war, perhaps nuclear war. The absolute number of cases of deterrence failure is, in itself, alarming. We regard it as intrinsically important to explain—and ultimately predict—cases of deterrence failure, irrespective of the general performance of the theory and the strategy. We contend that the essential first step is to trace the processes of deterrence failure in “critical” cases.⁴³

Our comparative analysis of critical cases of deterrence failure was designed to identify the conditions under which deterrence is likely to fail.

⁴¹ Achen and Snidal cite Ragin's persuasive argument that contextual and holistic social and political forces are frequently blurred and averaged in a meaningless way in “variable-oriented” approaches. Ragin, as does George, urges the use of case studies to trace detailed causal sequences in different contexts. Charles C. Ragin, *The Comparative Method* (Berkeley: University of California Press, 1987).

⁴² Edward A. Kolodziej, “The Limits of Deterrence Theory,” *Journal of Social Issues* 43, No. 4 (1987), 123-34, makes this point in the context of general deterrence.

⁴³ See Harry Eckstein, “Case Study and Theory in Political Science,” in Fred I. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science*, Vol. 7 (Reading, MA: Addison-Wesley, 1975), 79-138, and Arend Lijphart, “Comparative Politics and the Comparative Method,” *American Political Science Review* 65 (September 1971), 682-93, and “The Comparable Case-Strategy in Comparative Research,” *Comparative Political Studies* 8 (July 1975), 158-177, for a discussion of the importance of the analysis of “critical” cases.

In the analysis of these cases, we looked carefully at the relationships among variables specified by "rational deterrence theory."⁴⁴ We attempted to assess how and why differences in these relationships produced unanticipated outcomes and then to identify variables omitted by the theory and trace their interaction through the process of deterrence failure.⁴⁵ The empirical analysis of cases of deterrence failure consequently served two functions. It provided a partial test of "rational deterrence theory" insofar as it examined the interaction among variables the theory specifies. It also identified additional variables and their interrelationships that, our analysis demonstrated, have an important impact on the outcome of deterrence.

We explicitly acknowledge that our analyses provide only a partial test of theories of deterrence. Our analysis of deterrence failure can only be tentative, since some of the relationships among variables that appear to explain failure may also be at work when deterrence succeeds. Contingent propositions derived from a controlled comparison of deterrence failures must be validated against cases of deterrence success.

Deterrence succeeds when a deterrer explicitly attempts to prevent a use or deployment of military force and a would-be challenger considers initiating that action but is dissuaded by the defender's threat. These criteria of selection, we recognize, are likely to reduce the number of cases of immediate deterrence success we can identify. Since there are often no direct behavioral traces of success other than a defender's threat of retaliation, a search of the historical record is open-ended. Moreover, leaders are less likely to write and to speak, even privately, about an adversary's success in deterring them from a challenge. Identifying cases of general deterrence success over time is even more difficult. If deterrence succeeds over time, would-be challengers may never consider a challenge and defenders may never need to threaten explicitly or to manipulate the risk of war. In the absence of evidence, identification of the success of deterrence depends entirely on counterfactual argument. Achen and Snidal resort to precisely this kind of argument in their discussion of wars that did not occur.

Sensitive to the data requirements and to the threats to valid inference that arise from counterfactual argument, we are currently engaged in the identification of critical cases of immediate deterrence success. In all of

⁴⁴ See Lebow and Stein (fn. 11).

⁴⁵ For a discussion of the advantages of "process tracing" in critical cases, see Alexander L. George, "Case Studies and Theory Development," paper presented to the Second Annual Symposium on Information Processing in Organizations, Carnegie-Mellon University, Pittsburgh, PA, October 15-16, 1982.

these cases, we insist on convergent streams of evidence indicating that would-be initiators contemplated challenges but were dissuaded by deterrent threats. Evidence of the initiator's intentions and calculations is essential to determine whether the failure to challenge was the consequence of deterrence or of other considerations and constraints.

Once we have identified the critical cases for analysis, we intend to examine the propositions deduced from "rational deterrence theory" as well as those derived from the analysis of cases of deterrence failure. This comparative analysis will provide a more valid test both of existing theories of deterrence and the contingent propositions that have been identified through the empirical analysis of deterrence failure.

We contend that "rational deterrence theory" has not yet been tested satisfactorily. The first step has been taken. Analysis of deterrence failures has identified important variables omitted by the theory and traced their impact through to the failure of deterrence. Our analysis suggests that many, perhaps most, challengers are driven less by opportunity than by vulnerabilities and needs that motivate errors in their estimates of the variables specified by rational theories of deterrence as critical to the success of deterrence.

Comparative analysis of deterrence success is the essential next step if we are to assess the predictive power of "rational deterrence theory" as well as the validity of the contingent propositions derived from our analysis of deterrence failure. This analysis must be done with evidence from cases appropriate to the specifications of theory. With tongue only mildly in cheek, we might add that the responsibility for rigorous testing lies not only with the critics but also with the proponents of "rational deterrence theory." It is they, after all, who assert its predictive power as an empirical theory of strategic behavior. The challenge is not to assert but to demonstrate.

WHY DETERRENCE IS AN INADEQUATE THEORY

Our analysis has highlighted some of the more important conceptual failings of deterrence theory. We have argued that deterrence theories presuppose that leaders are (1) instrumentally rational, (2) risk-prone gain maximizers, (3) free of domestic constraints, and (4) able correctly to identify themselves as defenders or challengers. All of these core assumptions are unrealistic and contradicted by empirical evidence.

Achen and Snidal contend that this is not a problem for deterrence theory. They assert that the only criterion of good theory is its predictive capacity. They further contend that a theory based on unrealistic assump-

tions can still be a good predictor of behavior.⁴⁶ Philosophers of science challenge this assertion.⁴⁷ Although they disagree among themselves about the nature of social science theory and whether or not it can or ought to be scientific, they categorically reject the notion that good predictive theories can be built on unrealistic assumptions.

Theoretical models often contain two types of terms: those referring to unobserved but actual phenomena, and those referring to ideal entities. Neutrons, quarks, and black holes are examples of the former. They can be measured, counted, and detected through their interactions, and their existence and properties thereby confirmed or disconfirmed. Idealizations, by contrast, are derivative concepts that are not applicable to any empirical situation. Neither an ideal gas nor a perfect vacuum can be found in nature. They are concepts derived from other theoretical assumptions, fundamental to the model, that do have empirical referents. They serve an important function in theory: they provide extreme values against which empirical observations can be interpreted.

The problem with rational models is not that they contain idealizations but that these idealizations are their fundamental assumptions.⁴⁸ Such assumptions as the rational decision maker, perfect information, and a politically neutral environment are idealizations that lack any empirical referent. Rational deterrence theories are accordingly “theories” about nonexistent decision makers operating in nonexistent environments. If rational deterrence theories are to be useful as empirical theories, they must open their fundamental assumptions and behavioral implications to empirical evaluation, not protect or shield them from scrutiny.⁴⁹

⁴⁶ This is a line of argument that originated with Milton Friedman, “The Methodology of Positive Economics,” in Friedman, *Essays in Positive Economics* (Chicago: University of Chicago Press, 1953), 3-43. It entered political science through the work of Anthony Downs, *An Economic Theory of Democracy* (New York: Harper & Row, 1957).

⁴⁷ For an excellent discussion of the scientific problems of rational models, see Terry M. Moe, “On the Scientific Status of Rational Models,” *Journal of Political Science* 23 (February 1979), 215-43.

⁴⁸ *Ibid.*, 226.

⁴⁹ Moe makes this argument for rational models in general (*ibid.*, 235).