Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense

Robert Powell

On December 17, 2002, President George W. Bush ordered the deployment of a national missile defense (NMD). Proponents of missile defenses, both inside and outside the Bush administration, argue that, absent NMD, the proliferation of nuclear weapons and the greater U.S. vulnerability that this entails will significantly limit the United States´ ability to secure it foreign policy goals. "A policy of intentional vulnerability by the Western nations," Secretary of Defense Donald Rumsfeld argues, "could give rogue states the power to hold our people hostage to nuclear blackmail—in an effort to prevent us from projecting force to stop aggression."[1] Similarly, Walter Slocombe as undersecretary of defense in the Clinton administration asserted, "Without defenses, potential aggressors might think that the threat of strikes against U.S. cities could coerce the United States into failing to meet its commitments."[2]

To what extent do the spread of nuclear weapons and the means to deliver them threaten U.S. interests and impede the United States´ ability to pursue its

Robert Powell is Robson Professor in the Department of Political Science at the University of California, Berkeley.

I appreciate helpful discussion and criticisms from Jerome Bracken, Kevin Donovan, James Fearon, Charles Glaser, Avery Goldstein, Joanne Gowa, Lisbeth Gronlund, David Lake, Carter Malkasian, James Morrow, Dennis Powell, Robert Rauchhaus, Thomas Schelling, Leo Simon, Marc Trachtenberg, and those participating in the "Nuclear Policies Workshop," Institute on Global Cooperation and Conflict, University of California, San Diego, May 17–18, 2001; "Fourth Nuclear Stability Roundtable," Defense Threat Reduction Agency, Arlington, Virginia, March 12–13, 2002; and the "Weapons Study Forum," Los Alamos National Laboratory, May 23, 2002. I am also grateful for financial support from the National Science Foundation (SES-9911075).

1. Donald H. Rumsfeld, "Prepared Testimony to the Senate Armed Services Committee," June 21, 2001, p. 2, http://www.defenselink.mil/speeches/2001/s20010621-secdef.html.
2. Walter Slocombe, "The Administration´s Approach," *Washington Quarterly,* Vol. 23, No. 3 (Summer 2000), p. 79. Making similar claims are James M. Lindsay and Michael E. O´Hanlon, *Defending America: The Case for Limited Missile Defense* (Washington, D.C.: Brookings, 2001), p. 18; Keith B. Payne, "Post–Cold War Deterrence and Missile Defense," *Orbis,* Vol. 39, No. 2 (Spring 1995), pp. 201–223; and Paul Wolfowitz, "Prepared Testimony on Ballistic Missile Defense to the Senate Armed Services Committee," July, 12, 2001, http://www.defenselink.mil/speeches/2001/s20010712-depsecdef.html. These ideas are also incorporated in the December 2001 U.S. *Nuclear Posture Review* excerpts, which are available at http://www.globalsecurity.org/wmd/library/policy/dod/npr.html.

International Security, Vol. 27, No. 4 (Spring 2003), pp. 86–118

interests? To what degree would a national missile defense offset any adverse effects of this spread and at what political and economic cost? The answers to these questions depend at least in part on one´s ideas about how nuclear deterrence works.

Surprisingly though, most analyses of the effects of the spread of nuclear weapons and of the corresponding value of NMD have not grounded themselves directly in nuclear deterrence theory.[3] These studies have generally not drawn explicitly on the foundational work of Bernard Brodie, Hermann Kahn, Thomas Schelling, Glenn Snyder, Albert Wohlstetter, and a few others. Indeed some analyses have explicitly rejected nuclear deterrence theory, seeing it as an obsolete and possibly dangerous kind of Cold War thinking.[4] Other analyses, focusing more narrowly on specific policy or technical issues, have typically relied on a gloss on nuclear deterrence theory, variously asserting either that the threat of retaliation will deter rogue states as it deterred the Soviet Union during the Cold War or that this threat may prove insufficient to deter rogues.[5] And still other more theoretically oriented studies have looked to different theories and conceptual frameworks and not to nuclear deterrence theory.[6]

This article extends some of the fundamental ideas developed by Brodie, Snyder, and especially Schelling in the U.S.-Soviet context to examine the effects of the spread of nuclear weapons and of NMD. Three broad conclusions emerge. First, although nuclear deterrence theory remains useful, its implications vary with the conditions in which it is applied. Therefore, the relative stability between the United States and the Soviet Union during the second half

---

3. A notable exception is Avery Goldstein, *Deterrence and Security in the 21st Century* (Stanford, Calif.: Stanford University Press, 2000).
4. See, for example, Fred C. Iklé, "The Second Coming of the Nuclear Age," *Foreign Affairs,* Vol. 74, No. 1 (January–February 1996), pp. 119–128; and Keith B. Payne, *The Fallicies of Cold War Deterrence and a New Direction* (Lexington: University of Kentucky Press, 2001). This view also seems to be widely held in the Bush administration. See Steven E. Miller, "The Flawed Case for Missile Defense," *Survival,* Vol. 43, No. 3 (Autumn 2001), pp. 103–104.
5. Examples include George Lewis, Lisbeth Gronlund, and David Wright, "National Missile Defense: An Indefensible System," *Foreign Policy,* No. 117 (Winter 1999–2000), pp. 120–137; Ivo Daalder, James M. Goldgeier, and James M. Lindsay, "Deploying NMD: Not Whether, but How," *Survival,* Vol. 42, No. 1 (Spring 2000), pp. 6–28; Lindsay and O'Hanlon, *Defending America;* Spurgeon Keeny, "The New Missile 'Threat´ Gap," *Arms Control Today,* Vol. 28, No. 5 (June–July 1998), p. 2; and Walter B. Slocombe, "Stability Effects of Limited Missile Defenses: The Case for the Affirmative," *Pugwash Occasional Papers,* Vol. 3, No. 1 (March 2002), pp. 73–88.
6. Scott D. Sagan and Kenneth N. Waltz, *The Spread of Nuclear Weapons: A Debate* (New York: W.W. Norton, 1995), for example, offer one of the most theoretical analyses of the effects of the spread of nuclear weapons. But Waltz claims to derive his conclusions from a structural theory of international politics, whereas Sagan juxtaposes his organizational theory against Waltz´s formulation. See ibid., pp. 50–55, 112.

of the Cold War following the 1963 Cuban missile crisis may provide a poor guide to the stability of a crisis between the United States and a new nuclear state (or, for that matter, between two new nuclear states such as India and Pakistan). Second, NMD would give the United States somewhat more freedom of action and make a rogue state more likely to back down in a crisis. But these effects will be modest unless the defenses are very good. Finally, NMD, unless it is extremely effective, is likely to raise the risk both of a nuclear attack on the United States and of nuclear weapons striking the United States. These greater risks, moreover, are not the result of a mistaken overconfidence in the effectiveness of NMD. They are the direct consequence of a greater U.S. willingness to press its interests in a crisis harder.

There are three parts to the analysis. The first revisits and elaborates some of the elements of nuclear deterrence theory. It focuses especially on the credibility problem inherent in nuclear deterrence, the way that conflicts of interest play themselves out in the presence of nuclear weapons, and the dynamics of brinkmanship. It also shows that "rational deterrence theory," despite some claims to the contrary, actually helps to explain how deterrence between rational actors might fail and what makes this failure more likely. The second part discusses the dynamics of escalation and the likelihood of U.S. intervention if the United States is facing a rogue state and does not have any missile defenses. The last part examines the effects of NMD on stability and the likelihood of U.S. intervention.

Nuclear Deterrence Theory Revisited

In the 1950s and early 1960s, strategists and policymakers anticipated the arrival of a technological condition of mutual assured destruction in which both the United States and the Soviet Union could launch a devastating nuclear second strike even after absorbing a massive nuclear first strike. Secure, second-strike forces would render defense impossible as neither state could physically protect itself from an attack. Consequently, both states would have to rely on deterrence to dissuade the other from attacking should it be tempted to do so.[7]

7. Early contributions to deterrence theory include Bernard Brodie, ed., *The Absolute Weapon* (New York: Harcourt, 1946); Bernard Brodie, *Strategy in the Missile Age* (Princeton, N.J.: Princeton University Press, 1959); Hermann Kahn, *On Thermonuclear War* (Princeton, N.J.: Princeton University Press, 1960); Klaus Knorr and Thorton Read, eds., *Limited Strategic War* (Princeton, N.J.: Princeton University Press, 1962); Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, Mass.: Harvard University Press, 1960); Thomas C. Schelling, *Arms and Influence* (New Haven, Conn.: Yale Univer-

But this reliance would pose a profound theoretical challenge. The nuclear revolution, marked by the advent of secure, second-strike forces, would not eliminate political conflicts of interest. How would these conflicts play out in the face of these capabilities? How and to what extent could states exert coercive pressure on each other to further their interests—be those interests to protect what they already have or acquire more? A fundamental credibility problem lies at the heart of these questions, and one can see much of nuclear deterrence theory as an effort to resolve this problem.

A FUNDAMENTAL CREDIBILITY PROBLEM

A state´s assured-destruction capability gives it the ability to make the cost that an adversary has to bear in any conflict outweigh any possible gains. If, therefore, a state´s threat to impose these costs were sufficiently credible, an adversary would prefer backing down. Thus the ability to exert coercive pressure would seem to turn on the credibility of the threat. But how can a state credibly threaten to impose a sanction that, if imposed, would subsequently result in its own destruction? Indeed, given that both states have second-strike capabilities and can therefore make the costs outweigh any gains, why would either state be any more able to exert coercive pressure on its adversary than its adversary would be able to exert on it? Why do these capabilities not simply cancel each other out?

Thomas Schelling devised a solution to this problem or, more precisely, a way around it. Because neither state can physically protect itself from an adversary´s attack, the probability of escalation to catastrophic levels of mutual destruction is always present in any crisis. Indeed "it is the essence of a crisis that the participants are not fully in control of events."[8]

The risk of accidental or inadvertent escalation to nuclear war is the key to solving the credibility problem and to understanding the dynamics of coercion. In a condition of mutual assured destruction, states cannot credibly threaten to launch a massive nuclear attack deliberately. But they may be able to credibly make "threats that leave something to chance."[9]

---

sity Press, 1966); Glenn H. Snyder, *Deterrence and Defense* (Princeton, N.J.: Princeton University Press, 1961); and Albert Wohlstetter, "The Delicate Balance of Terror," *Foreign Affairs,* Vol. 37, No. 2 (January 1959), pp. 211–234. "Strategic Thought in America, 1952–1966," in Marc Trachtenberg, *History and Strategy* (Princeton, N.J.: Princeton University Press, 1991), pp. 3–46, reviews the development of these ideas. Robert Jervis, *The Illogic of American Nuclear Strategy* (Ithaca, N.Y.: Cornell University Press, 1984); and Robert Jervis, *The Meaning of the Nuclear Revolution* (Ithaca, N.Y: Cornell University Press, 1989), provide important subsequent elaboration and extensions.

8. Schelling, *Arms and Influence,* p. 97.
9. Schelling, *The Strategy of Conflict,* p. 187.

That is, a state may be able to credibly threaten and actually engage in a process—a crisis or a limited war—that raises the risk that the situation will go out of control and escalate to a catastrophic nuclear exchange. How much risk a state could credibly threaten to run would depend on what was at stake in the political conflict. The higher the stakes, the more risk a state could credibly threaten to run.[10]

Crises in this view become a kind of brinkmanship. During a crisis, states exert coercive pressure on each other by taking steps that raise the risk that events will go out of control. This is a real and shared risk that the confrontation will end in a catastrophic nuclear exchange. Consequently, no state bids up the risk eagerly or enthusiastically. Rather a state faces a series of terrible choices throughout the conflict. A state can quit, or it can decide to hang on a little longer and accept a somewhat greater risk in the hope that its adversary will find the situation too dangerous and back down. If neither state backs down, the crisis goes on with each state bidding up the risk until one of the states eventually finds the risk too high and backs down or until events actually do spiral out of control.[11]

Brinkmanship clearly does not imply that states engage in bold or reckless behavior. States may be very reticent to raise the risk. Nevertheless, this point has often been misunderstood, and the fact that states usually are very reluctant to run risks has sometimes been cited as evidence against the brinkmanship model of crisis bargaining. For example, Richard Betts, in his compre-

10. Schelling´s threats that leave something to chance solved the credibility problem conceptually, and he drew on what was then known about the Cuban missile crisis to illustrate the idea. See Schelling, *Arms and Influence,* pp. 94–97. But Schelling´s work was primarily theoretical, and he did not offer any systematic empirical analysis showing that events could really go out of control in the ways postulated in these threats. Subsequent work on command and control and the risk of accident and inadvertence has done much to fill this gap. See, for example, Bruce G. Blair, *The Logic of Accidental Nuclear War* (Washington, D.C.: Brookings, 1993); Bruce G. Blair, "Nuclear Inadvertence: Theory and Evidence," *Security Studies,* Vol. 3, No. 3 (Spring 1994), pp. 494–500; Paul J. Bracken, *The Command and Control of Nuclear Forces* (New Haven, Conn.: Yale University Press, 1983); Peter D. Feaver, *Guarding the Guardians: Civilian Control of Nuclear Weapons in the United States* (Ithaca, N.Y.: Cornell University Press, 1992); Peter D. Feaver, "The Politics of Inadvertence," *Security Studies,* Vol. 3, No. 3 (Spring 1994), pp. 501–508; Scott D. Sagan, *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons* (Princeton, N.J.: Princeton University Press, 1993); and Scott D. Sagan, "Organized for Accidents," ibid., pp. 509–520. For a review of this work, see Bradley A. Thayer, "The Risk of Nuclear Inadvertence," ibid., pp. 428–493.

11. Schelling, *The Strategy of Conflict,* pp. 187–203; and Schelling, *Arms and Influence,* pp. 92–125. Relying on the risk of events going out of control was one of two broad approaches to solving the credibility problem. The second approach, which received much less attention, was based on the idea of limited, countervalue attacks. See Knorr and Read, *Limited Strategic War,* for a discussion of this approach; and Robert Powell, "The Theoretical Foundations of Nuclear Deterrence," *Political Science Quarterly,* Vol. 100, No. 1 (Spring 1985), pp. 75–96, for a comparison of these two approaches.

hensive study of nuclear crises, describes the strategy based on Schelling´s threats that leave something to chance as a "risk-maximizing approach" in which a state "tends toward *maximizing* and accepting mutual military risk."[12] His empirical findings, however, are generally consistent with brinkmanship.

Marc Trachtenberg also believes that brinkmanship implies that statesmen will be eager to run risks in order to out bid the other side.[13] Nevertheless, his assessment of what happened during the Cuban missile crisis is in keeping with the formulation above in which states are very reluctant to press on and raise the risk, but even more reluctant to back down. He concludes that the evidence shows that: "1) leading officials believed that nuclear war could come without either side having to make a cold-blooded decision to start one; 2) these officials were willing during the crisis to accept a certain risk of nuclear war; and 3) the risk of war was consciously manipulated in order to affect Soviet options in the crisis."[14]

In sum, brinkmanship provides a model of the way that states can exert coercive pressure on each other if both have secure, second-strike capabilities. Like all models, it simplifies and abstracts by emphasizing some things while minimizing or ignoring others. The value of the model lies in the extent to which it explains and illuminates the dynamics of escalation.

THE DYNAMICS OF BRINKMANSHIP

Brinkmanship is fundamentally a contest of resolve in which states bid up the risk of events spiraling out of control until one of the states finds this risk intolerably high and backs down. How does this contest play out? What determines the dynamics of brinkmanship? As is shown below, there are no crises if there is little or no uncertainty about the states´ levels of resolve. In this case, the less resolute state does not challenge the more resolute state. Crises arise only if there is substantial uncertainty about the balance of resolve, and in this case, the dynamics of escalation depend on a complex interaction between the states´ levels of resolve and their uncertainty about each other´s resolve.

The first step in deriving these conclusions and illustrating this complicated

12. Richard K. Betts, *Nuclear Blackmail and Nuclear Balance* (Washington, D.C.: Brookings, 1987), pp. 10–11 (emphasis in the original).

13. Marc Trachtenberg, "The Influence of Nuclear Weapons in the Cuban Missile Crisis," *International Security,* Vol. 10, No. 1 (Summer 1985), p. 146.

14. Ibid., p. 140. For additional examples, see Robert Jervis, "Deterrence Theory Revisited," *World Politics,* Vol. 31, No. 2 (January 1979), pp. 303–304; and Robert Jervis, "The Political Effects of Nuclear Weapons: A Comment," *International Security,* Vol. 13, No. 2 (Fall 1988), p. 80; as well as Paul K. Huth, "Deterrence and International Conflict," *Annual Review of Political Science,* Vol. 2 (1999), p. 38.

interaction is to describe the notion of resolve more precisely. During a brinkmanship crisis, a state faces a choice between acquiescing and continuing to run the risk that events will go out of control and result in a nuclear exchange. Define a state´s resolve, *R*, to be the maximum risk of suffering a nuclear strike that the state is willing to run in order to prevail.[15] This definition means that a state would acquiesce and the crisis would end if the state (somehow) knew that it would have to run a risk larger than *R* to prevail. By contrast, a state would continue on in a crisis if it (somehow) knew that the other state would quit before the risk exceeded *R*.[16]

Resolve is the avenue through which the political conflict underlying the crisis enters the analysis and affects the dynamics of escalation. Resolve is the reason why two adversaries´ nuclear forces do not simply cancel each other out in brinkmanship, and why one side may have an advantage over the other. The more a state values prevailing, the more risk it would be willing to run in order to do so and the greater its resolve. Similarly, the higher the cost of acquiescing, the more risk it would be willing to tolerate in order to avoid losing and the greater its resolve. And the more catastrophic the outcome if events go out of control, the less risk a state would be willing to accept and the lower its resolve.

To see how the balance of resolve and the uncertainty surrounding it affect the dynamics of brinkmanship, consider the very artificial situation in which there is no uncertainty about the states´ levels of resolve. Each state knows its resolve and that of the other state. The dynamics of escalation are very simple in this case: There are none. Absent uncertainty, the less resolute state knows that no matter how hard it pushes a crisis, the more resolute state would be willing to escalate still further. Indeed if the more resolute state has to, it is willing to push the crisis to the point where the risk exceeds the less resolute state´s resolve. At this point, the situation becomes too dangerous for the less resolute state and it backs down. Given that the less resolute state cannot prevail, the choice it faces is between backing down at the very outset before generating any risk and backing down later if events have not gone out of control

15. There are of course many different kinds of nuclear attack, and I am simplifying here. This simplification is discussed further below.
16. For more extensive discussions of resolve, see Robert Jervis, "Bargaining and Bargaining Tactics," in James Roland Pennock and John W. Chapman, eds., *Coercion* (Chicago: Aldine-Atherton, 1972); Glenn H. Snyder and Paul Diesing, *Conflict among Nations* (Princeton, N.J.: Princeton University Press, 1977); and Robert Powell, "Crisis Bargaining, Escalation, and MAD," *American Political Science Review,* Vol. 81, No. 3 (September 1987), pp. 717–736.

in the interim. The first alternative is clearly better. So the less resolute state never escalates if there is no uncertainty about the states´ levels of resolve.

The same dynamics obtain in the somewhat less artificial situation in which neither state knows the exact level of the other´s resolve but each knows who is more resolute. As long as the balance of resolve is clear, the less resolute state still knows that the more resolute state is willing to push the crisis harder and run more risk. The less resolute state therefore never escalates, and there are no brinkmanship crises.

Brinkmanship crises occur only if the balance of resolve is uncertain. If each state believes that it is likely to be more resolute than the other state, then each may escalate in the expectation that the other will back down. As the crisis continues and neither state backs down, each learns that the other is more resolute than it initially believed. Eventually one state concludes that the risk is too high and the chances that the other will back down are too low to warrant further escalation. At that point, that state backs down and the crisis ends—assuming of course that events have not already gone out of control.

How much risk do the states run, and which state backs down first? The following game-theoretic model of brinkmanship helps to sort through the complicated interaction between the states´ levels of resolve and the uncertainty surrounding them. In the simplest version of brinkmanship, the longer the crisis goes on, the higher the risk that it will go out of control. In this setting, each state simply decides how long it will let the crisis go on and, therefore, how high it would be willing to let the risk go.

To fix ideas, let $p(t)$ denote the probability that the crisis goes out of control by time t. Suppose further that state S_1 decides that it is willing to let the crisis continue until time t_1 or equivalently until the risk of disaster is $p(t_1)$. S_2 decides that it is willing to let the crisis go until t_2 and risk $p(t_2)$ where $t_2 > t_1$. Because S_2 is willing to hold on longer than S_1 ($t_2 > t_1$), the crisis lasts until it goes out of control at some time before t_1 or until S_1 acquiesces at t_1. If the crisis lasts until S_1 quits at t_1, then the states will have run a risk of $p(t_1)$ of events spiraling out of control.[17]

17. Crises in this simplest formulation are essentially generators of risk over which the states´ only influence is to decide when to stop. The states cannot, for example, affect the tempo of the crisis by taking steps that would increase or decrease the rate at which the risk changes. More formally, the states have no influence over the shape of the function $p(t)$. Allowing the states more influence would complicate the analytics enormously and is beyond the scope of this article. But pursuing this in future work might help resolve a puzzle in nuclear deterrence theory. Intuitively, a state that expected to prevail at a given level of conflict or rung of the escalation ladder would like to reduce the risk of events going out of control. This line of reasoning might provide a theoretical founda-

How high a state is willing to let the risk go before backing down depends on its resolve and on its uncertainty about the other state´s resolve. Accordingly, the first step in determining how high a state will allow the risk to go is to specify the states´ payoffs and formalize the notion of resolve. To this end, observe that the crisis can conclude in this simple model in one of only three ways: A state backs down, it prevails because the other backs down, or the crisis goes out of control. Simplifying still further, assume that the payoffs to these outcomes remain constant throughout the crisis with w_1 denoting S_1's payoff to prevailing, s_1 its cost to submitting, and d_1 its cost if events go out of control.[18] S_2's payoffs are defined analogously.

Given these payoffs, S_1 would be willing to run a given risk of a nuclear strike in order to prevail if its expected payoff to doing so is at least as great as its payoff to backing down. S_1's resolve, R_1, is the largest risk that it would be willing to run. In symbols, this largest risk is $R_1 = (w_1 + s_1)/(w_1 + d_1)$. S_2's resolve is defined in a parallel fashion.[19] This specification of resolve formalizes the role played by the political stakes underlying the crisis. A state´s resolve increases as its payoff to prevailing or cost to submitting go up, or its cost of events going out of control decreases.

The brinkmanship game is specified formally in the appendix. Less formally, each state decides how long it is willing to hang on given its own level of resolve and its beliefs about the other´s resolve. Because the risk that a crisis will go out of control rises as the crisis continues, deciding how long to hang on is equivalent to deciding how high a state is willing to let the risk go before quitting. Let $r_1(R_1)$ denote how high S_1 is willing to bid up the risk given that its resolve is R_1, and let $r_2(R_2)$ denote the same for the other state.

These bids determine how much risk the states run, which state prevails if the crisis does not go out of control, and what their payoffs are. More specifically, the state that bids the most risk (i.e., is willing to hang on longer)

---

tion for "escalation dominance," which some analysts argued during the Cold War was inconsistent with the logic of the nuclear revolution. See Robert Jervis, *The Illogic of American Nuclear Strategy* (Ithaca, N.Y.: Cornell University Press, 1984), especially pp. 59–63.

18. Allowing the state´s payoffs to vary as the crisis continues makes the game much more difficult to analyze. Because prevailing is better than submitting and submitting is better than a catastrophic nuclear exchange, $w_1 \geq 0 \geq -s_1 \geq -d_1$ (where the payoffs to the latter two outcomes are the negative of the costs of these outcomes).

19. To derive this expression, observe that S_1's expected payoff to running a risk of disaster r is just its payoff to prevailing, weighted by the probability of avoiding attack, less the cost to disaster, weighted by the probability of that outcome. This is $w_1(1 - r) - d_1r$. Consequently, S_1 would be willing to hazard a risk of r if $w_1(1 - r) - d_1r \geq -s_1$, and the largest risk for which this relation holds is R_1 as defined above.

prevails but has to run the risk determined by how long the other state hung on. If, for example, S_1's bid is higher—that is, if $r_1(R_1) > r_2(R_2)$, then S_1 prevails but has to run risk $r_2(R_2)$. If both states bid the same risk, neither backs down and the situation on the ground determines the outcome.[20]

Brinkmanship is in effect a kind of auction.[21] In a typical or so-called English auction, the bidders bid up the price until no one is willing to bid more than the last bid. At that point, the auction ends, the highest bidder gets the item on offer for the price she bid, and no one else pays anything. This, however, is only one of many different types of auction. In a second-price auction, for example, the highest bidder wins but pays the second-highest bid. In yet another type of auction—called an all-pay auction—the highest bidder still wins, but everyone pays the price that this individual bid.[22]

Brinkmanship can be seen as a variant of an all-pay, second-price auction in which bids are measured, not in terms of money, but in terms of the risk that events will go out of control. During a crisis, each state bids up the risk until one of the states finds the risk too high and quits. The state that prevails is the one that is willing to hang on longer, that is, makes the highest bid. But the amount of risk that the states actually run during the crisis is determined by the state that backs down first. Thus, the "price" that each state must pay—that is, the risk of disaster each must run—is determined not by the highest bid but by the second-highest bid. (Recall that if S_1 bids r_1 and S_2 bids a lower risk r_2, then the crisis continues until the risk reaches the lower level r_2 at which point S_2 quits.) This makes brinkmanship an all-pay, second-price auction.

Thinking about brinkmanship in this way highlights the critical role that uncertainty about the balance of resolve plays. This formulation also makes it clearer why there are no crises if the balance of resolve is unambiguous. In these circumstances, the less resolute state knows that the more resolute state will always outbid it. This implies that the less resolute state cannot win the auction. But because this is an all-pay auction, bidding is costly to everyone,

20. Barry Nalebuff, "Brinkmanship and Nuclear Deterrence: The Neutrality of Escalation," *Conflict Management and Peace Science,* Vol. 9, No. 2 (Spring 1986), pp. 19–30, shows that the risk the states run (in equilibrium) does not depend on the shape of the function p. If, for example, this function is steeper so that crises become more dangerous more quickly, states offset this greater danger by dropping out sooner. Intuitively, states care about risks, not time. The formulation here models this more directly by letting the states decide directly on the risks $r_1(R_1)$ and $r_2(R_2)$.
21. See Powell, *Nuclear Deterrence Theory,* pp. 154–155.
22. For an introduction to auction theory, see R. Preston McAfee and John McMillian, "Auctions and Bidding," *Journal of Economic Literature,* Vol. 25, No. 2 (June 1987), pp. 699–738; and Paul Milgrom, "Auctions and Bidding: A Primer," *Journal of Economic Perspectives,* Vol. 3, No. 3 (Summer 1989), pp. 3–22.

not just the eventual winner. If, therefore, a state knows that it cannot win and will have to pay even if it bids in a losing effort, it will not bid.

When the balance of resolve is uncertain, a state faces a trade-off when deciding how much to bid. The more risk it is willing to accept, the more likely the other side is to back down first. But the more it bids, the more likely events are to go out of control. The optimal bid equates the marginal gain to holding out slightly longer with the marginal cost of doing so. A solution or equilibrium of the brinkmanship game is a pair of bidding strategies $(r_1(R_1), r_2(R_2))$ such that each state´s bid maximizes its payoff given its level of resolve, its beliefs about the other state´s level of resolve, and the other state´s strategy.

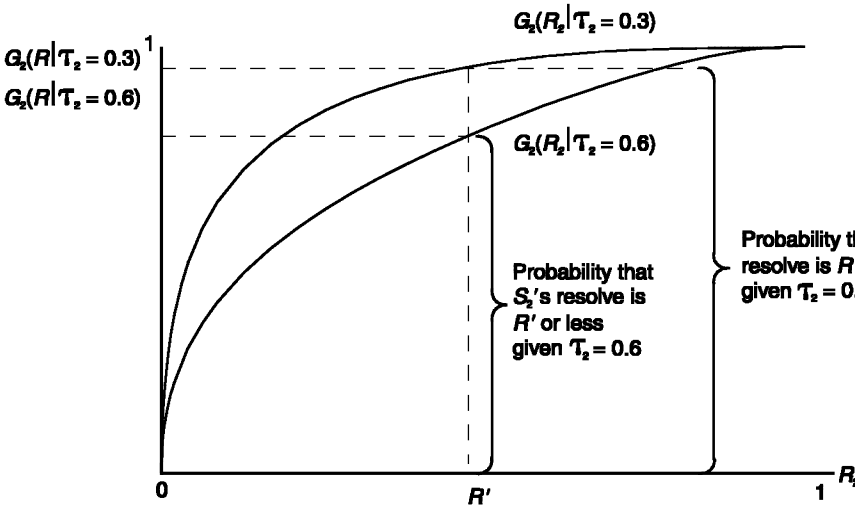
In game-theoretic models, actors´ beliefs are represented by probability distributions. In the present game, each state is uncertain of the other´s level of resolve, and S_1´s beliefs about S_2´s resolve is represented by the probability distribution $G_2(R_2 \mid \tau_2)$. The parameter τ_2 indicates how "tough" S_1 thinks S_2 is. As τ_2 increases, S_1 becomes increasingly confident that S_2 is highly resolute.

Figure 1 illustrates the effects of changes in τ_2 and how these changes capture the idea of S_2´s toughness. Suppose, for example, that τ_2 equals 0.3. Then S_1 believes that the probability that S_2´s resolve is, say, R' or less is $G_2(R' \mid \tau_2 = 0.3)$. If, by contrast, $\tau_2 = 0.6$, then S_1 believes that the chances that S_2´s resolve is R' or less is $G_2(R' \mid \tau_2 = 0.6)$. To see why S_2 seems tougher when τ_2 is larger, consider S_1's beliefs about how likely its adversary´s resolve is to be above any given level. The more likely an adversary´s resolve is to be above a given level, the more resolute that adversary seems. And the probability that S_2´s resolve is above any given level, say R', rises as τ_2 increases. Hence, S_2 seems tougher to S_1 the larger τ_2.[23]

The functions defining the states´ equilibrium bids $(r_1(R_1), r_2(R_2))$ are derived in the appendix. These functions show how a state´s bid varies with its resolve, its uncertainty about the other state´s resolve, and the other state´s uncertainty about its resolve. Not surprisingly, the more resolute a state is, the harder it is willing to press the crisis. Further, the tougher or more resolute a state believes its adversary to be (i.e., the higher the other state´s τ), the less willing it is to press the crisis. By contrast, the more resolute a state is believed to be (i.e., the higher a state´s own τ), the more it can exploit this perceived toughness to its advantage by being willing to push the crisis somewhat harder.[24] All of this

23. In symbols, the probability that S_2's resolve is at least as large as R' is $1 - G_2(R' \mid \tau_2)$. This probability increases as τ_2 increases because $G_2(R' \mid \tau_2)$ goes down as τ_2 rises. S_2's beliefs about S_1's resolve, $G_1(R \mid \tau_1)$, are defined analogously.
24. More formally, $r_1(R_1)$ is increasing in R_1 and τ_1, decreasing in τ_2, with analogous results for $r_2(R_2)$.

Figure 1. S_1's Beliefs about S_2's Resolve.



implies that the more resolute a state is believed to be, the higher its expected payoff; and the more resolute a state believes its adversary to be, the lower that state's payoff.

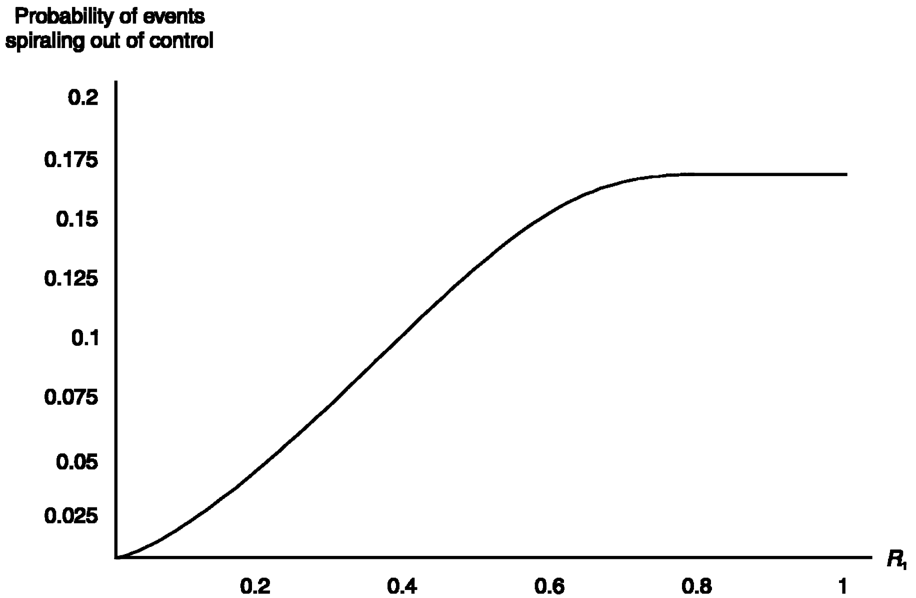
STABILITY

What does all of this say about stability—that is, about the probability that a crisis ultimately goes out of control and ends in a nuclear exchange? How, more specifically, does stability vary with the states' levels of resolve and their uncertainty about the balance of resolve? As just noted, the more resolute a state is, the longer it is willing to hang on and the more risk it is willing to run before backing down. This makes a crisis more dangerous. Figure 2 traces this effect by plotting the risk that S_1 runs in a crisis as a function of its resolve in a "nominal" case in which each state believes that there is a 50-50 chance that the other's resolve is 10 percent or less.[25] As the figure shows, the greater S_1's resolve, the more likely the crisis is to end with events spiraling out of control.[26]

25. A 50-50 chance that each state's resolve is 10 percent or less means that τ_2 and τ_2 are approximately 0.5. The parameter values in this and subsequent examples are only meant to be suggestive, and the parameters for all of the figures are given in Table 1 in the appendix.

26. The risk of events going out of control increases very slowly for high values of R_1, say 0.7 and above, because S_2 is very likely to have quit by the time the risk reaches the level bid by S_1 with $R_1 = 0.7$. Hence, S_1's willingness to hang on longer and bid more than this does not entail much additional risk.

Figure 2. The Effects of Resolve on Stability.

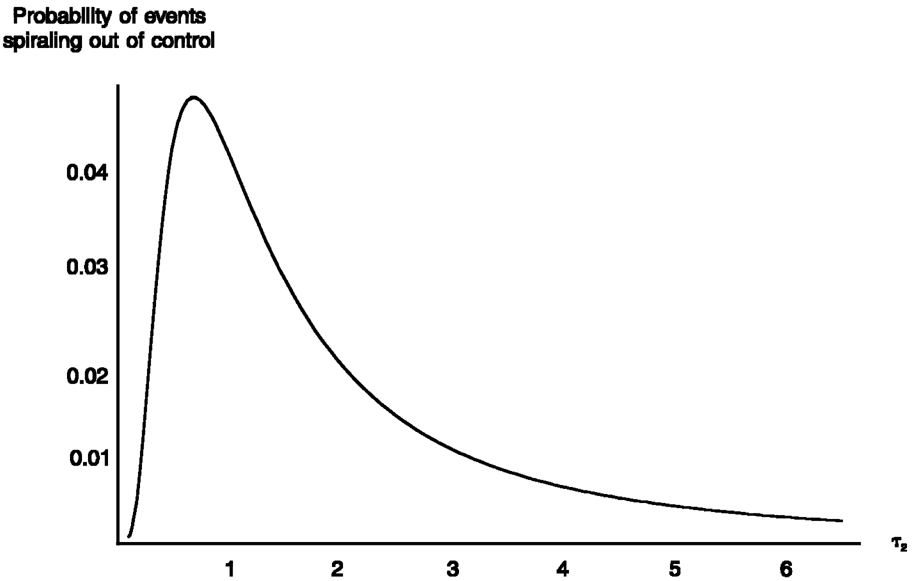


Observe, however, that the risk that a state actually runs is generally quite low relative to its resolve. For example, S_1 runs a risk of about 1.5 percent if its resolve is 10 percent.[27]

Figure 3 illustrates the effects of uncertainty about the balance of resolve on stability. Recall that τ_2 measures S_1´s beliefs about S_2´s resolve. At low values of τ_2, S_1 believes that the other state´s resolve is very likely to be quite low. Consequently, the balance of resolve in these circumstances is relatively clear (and in S_1´s favor), and the states run little risk in brinkmanship. As τ_2 begins to rise, S_1, still unsure of its adversary´s exact level of resolve, believes that it is higher on average. As a result, S_1 becomes less confident that it is more resolute than the other state; the balance of resolve begins to blur; and the risk of events ulti-

27. It might seem that S_1 should raise its bid if its resolve is 10 percent and it is running only a 1.5 percent chance of events going out of control. But recall that a state´s resolve is the maximum risk it would be willing to run given that it is sure to prevail if it runs that risk and if events do not spiral out of control. Bidding more than 1.5 percent in the present circumstances, however, does not bring the certainty of prevailing. It brings only a somewhat higher probability that S_2 will back down and that S_1 will prevail. But this better chance of prevailing comes at the cost of a somewhat higher chance of events going out of control. S_1's optimal bid balances the marginal gain to bidding more with the marginal cost of doing so, and this optimal bid generally entails running a risk much less than a state´s actual level of resolve.

Figure 3. The Effects of Uncertainty on Stability.



mately going out of control begins to rise. As τ_2 continues to rise, S_2 becomes increasingly confident that it is less resolute than its adversary. The balance of resolve becomes clearer, and the risk of events going out of control starts to decrease.

Brinkmanship and threats that leave something to chance offer a solution to the credibility problem at the center of nuclear deterrence theory. They provide a model of the way that states can exert coercive pressure on each other and that political conflicts can play out in situations in which no state can credibly threaten deliberately to trigger a catastrophic nuclear exchange. Stability in turn depends critically on the clarity of the balance of resolve. If the balance is clear, there will be no crisis because the less resolute state will not resist the more resolute state. If the balance of resolve is uncertain, each state may believe that the balance of resolve favors it and may escalate in the expectation that the other state will back down. Indeed brinkmanship—a kind of rational deterrence theory—predicts that there is some chance that deterrence will fail when the balance of resolve is sufficiently uncertain as Figure 3 illustrates.[28]

28. Sagan labels Waltz´s argument that the probability of major war among nuclear-armed states is almost zero "rational deterrence theory." Sagan and Waltz, *The Spread of Nuclear Weapons*, p. 51. This label is unfortunate. As Christopher H. Achen and Duncan Snidal emphasized at the outset of

There is, however, a critical distinction between a theory and the predictions derived from it. The latter depend not only on the theory but on the initial conditions as well. Who wins an auction and at what price depend not only on the rules of the auction (e.g., first-price, second-price, or all-pay) but also on how much the bidders value what is being auctioned off and on their uncertainty about the other bidders´ valuations. As argued below, the brinkmanship model can be applied in the post–Cold War context. But this does not mean that the United States will be able to deter its adversaries as well as it arguably did during the Cold War. Who prevails and at what risk in brinkmanship depends on the balance of resolve and the clarity of that balance. Should these factors differ in post–Cold War conflicts from what they were during the Cold War, then the Cold War experience will provide a poor guide to the future.[29] The remainder of this article examines how proliferation and NMD are likely to affect the balance of resolve and the degree of uncertainty surrounding it in a confrontation between the United States and a small nuclear state.

Brinkmanship and Rogue States

This section argues that brinkmanship can be used to analyze a confrontation between the United States and a small nuclear state—whether or not the latter is a rogue. It then discusses what it means to be a rogue in the context of the brinkmanship model and analyzes the dynamics of escalation. The analysis shows that when the balance of resolve clearly favors a small nuclear state, that state will be able to deter the United States. In particular, the United States will

---

the latest wave of the rational deterrence debate, this theory is about explaining why deterrence among rational actors sometimes succeeds and sometimes fails. Achen and Duncan Snidal, "Rational Deterrence Theory and Comparative Case Studies," *World Politics,* Vol. 41, No. 2 (January 1989), p. 152. For other fully rational (game-theoretic) models in which nuclear deterrence may fail, see Nalebuff, "Brinkmanship and Nuclear Deterrence"; and Robert Powell, *Nuclear Deterrence Theory: The Search for Credibility* (New York: Cambridge University Press, 1990). James D. Fearon, "Rationalist Theories of War," *International Organization,* Vol. 49, No. 2 (Summer 1995), pp. 379–414, offers an excellent overview of nonnuclear rationalist theories of war.

29. As Richard K. Betts observes, the relatively sanguine view that some take about the way deterrence worked during the Cold War is largely based on the second half of the Cold War following the Cuban missile crisis. But international politics seemed much more dangerous and fraught with peril during the first part of the Cold War, when the balance of resolve over various issues was much less clear. According to Betts, "The early phase of the Cold War, before the crises over Berlin and Cuba worked out the limits to probes and provocations, is a less reassuring model. Only with hindsight is it easy to assume that because the superpowers did not go over the edge, it was foreordained by deterrence that they could not." Betts, "Universal Deterrence or Conceptual Collapse? Liberal Pessimism and Utopian Realism," in Victor A. Utgoff , ed., *The Coming Crisis: Nuclear Proliferation, U.S. Interests, and World Order* (Cambridge, Mass.: MIT Press, 2000), p. 71.

generally be deterred from trying to overthrow the regimes of smaller nuclear states. Over issues where the balance of resolve is more ambiguous, the higher the risk the smaller nuclear state is willing to run—and in this sense the more of a rogue that state is—the more likely the United States is to be deterred from opposing or intervening against that state. Furthermore, a rogue is less likely to back down and more likely to push a crisis forward should the United States intervene.

Brinkmanship may provide a model for the way that political conflicts play out when two conditions hold. First, there is an outcome that both states view as being worse than giving in to the other state in the dispute. This creates the fundamental credibility problem discussed earlier, because neither state can credibly threaten to initiate a course of action that is certain to lead to this outcome. But, second, states can make threats that leave something to chance. That is, the states′ efforts to further their interests—perhaps through the use of military force—raise the risk that events will go out of control and end in the worse outcome. Both of these conditions appear to hold in a confrontation between the United States and a small nuclear state.

U.S. policymakers are likely to view the death and destruction wrought by even a very limited nuclear attack on the United States as being far worse than the consequences of not intervening against a regional nuclear power. During the Cold War, McGeorge Bundy argued that a decision that led to an attack resulting in even one bomb on one city would be a "catastrophic blunder."[30] Steve Fetter calculates that a 20 kiloton nuclear weapon dropped on a sparsely populated city lacking civil defense would kill 40,000 people and injure 40,000 more.[31] As Charles Glaser and Fetter observe, U.S. interests in regional disputes generally "are not truly vital, making it hard to justify pursuing foreign policies that increase the probability of attacks with weapons of mass destruction against U.S. cities."[32] Indeed this is precisely why some proponents of NMD favor its deployment. As the *Nuclear Posture Review* put the point in December 2001, missile defenses "can bring into better balance U.S. stakes and risks in a regional confrontation."[33]

Would the leaders of new nuclear states rather endure the overwhelm-

30. McGeorge Bundy, "To Cap the Volcano," *Foreign Affairs,* Vol. 48, No. 1 (October 1969), p. 10.
31. Steve Fetter, "Ballistic Missiles and Weapons of Mass Destruction: What Is the Threat? What Should be Done?" *International Security,* Vol. 16, No. 1 (Summer 1991), p. 27.
32. Charles L. Glaser and Steve Fetter, "National Missile Defense and the Future of U.S. Nuclear Weapons Policy," *International Security,* Vol. 26, No. 1 (Summer 2001), p. 69.
33. See http://www.globalsecurity.org/wmd/library/policy/dod/npr.htm, p. 4.

ing retaliatory response triggered by a nuclear attack on the United States than back down in a confrontation?[34] If so, then these states actually would be "undeterable." Rogue states are sometimes described as literally undeterable.[35] But there is little evidence for this, and as developed more fully below, a more careful formulation of the notion of a rogue state is that it is willing to run a greater risk of suffering this retaliatory outcome than another state would be. So, the first condition needed to apply brinkmanship seems to hold.

The second needed condition is that states´ actions can affect the risk that events will go out of control through accident or inadvertence. This risk has been the subject of much recent scholarship. Bruce Blair, Peter Feaver, Scott Sagan, Bradley Thayer, and others argue that the command-and-control problems facing emerging nuclear states will be more severe and that the risk of accidental or inadvertent war will be higher than it was for the United States and the Soviet Union.[36] By contrast, David Karl and Jordan Seng argue that the inherent risk of accidental or inadvertent use will be smaller for emerging nuclear states, but neither claims that this smaller risk is unimportant.[37] Most prominently, Kenneth Waltz argues that there is no reason to believe that the risk should be higher for emerging nuclear states. "Why should we expect new nuclear states to experience greater difficulties [with command and control]

34. Even those who question the wisdom or credibility of a nuclear response to a chemical or biological attack against the United States generally do not question this response to a nuclear attack. See, for example, Scott D. Sagan, "The Commitment Trap: Why the United States Should Not Use Nuclear Threats to Deter Biological and Chemical Weapons Attacks," *International Security*, Vol. 24, No. 4 (Spring 2000), p. 88.
35. Robert S. Litwak, *Rogue States and U.S. Foreign Policy: Containment after the Cold War* (Washington, D.C.: Woodrow Wilson Center Press, 2000), traces the conceptual and political origins and development of the notion of rogue states and U.S. policy toward them. See also Stephen M. Walt, "Containing Rogues and Renegades," in Utgoff, *The Coming Crisis*, pp. 194–199.
36. Blair, *The Logic of Accidental Nuclear War;* Blair, "Nuclear Inadvertence"; Feaver, *Guarding the Guardians;* Feaver, "The Politics of Inadvertence"; Sagan, *The Limits of Safety;* Sagan, "Organized for Accidents"; and Thayer, "The Risk of Nuclear Inadvertence." Peter R. Lavoy, "The Strategic Consequences of Nuclear Proliferation," *Security Studies*, Vol. 4, No. 4 (Summer 1995), pp. 695–753, also discusses these issues.
37. David J. Karl, "Proliferation Pessimism and Emerging Nuclear Powers," *International Security*, Vol. 21, No. 3 (Winter 1996/97), pp. 87–119, especially, pp. 113–114. Jordan Seng, "Less Is More: Command and Control Advantages of Minor Nuclear States," *Security Studies*, Vol. 6, No. 4 (Summer 1997), pp. 50–92; and Jordan Seng, "Optimism in the Balance: A Response to Peter Feaver," *Security Studies*, Vol. 6, No. 4 (Summer 1997), pp. 126–136, especially, p. 125. For a response to Karl and Seng, see Peter D. Feaver, "Neooptimists and the Enduring Problem of Nuclear Proliferation," *Security Studies*, Vol. 6, No. 4 (Summer 1997), pp. 93–125. Unlike Seng and to a somewhat lesser extent Karl, most others who argue that deterrence will work between emerging nuclear states do not focus directly on command-and-control issues and are not cited here. For references to that work, see Seng, "Less Is More," p. 51, n. 2.

than the ones old nuclear states were able to cope with," given that these new nuclear states have every incentive to control these risks?[38] Nevertheless, Waltz believes the risk that nuclear weapons might be used, albeit possibly quite small, is what exerts the deterrent pressure.[39] Implicitly then, this risk cannot be zero and must be large enough to influence behavior should deterrence fail and the states begin to press a crisis.

In sum, the two conditions necessary for brinkmanship would seem to hold in a confrontation between the United States and a regional nuclear state. The remainder of this section examines this confrontation through the lens of brinkmanship.

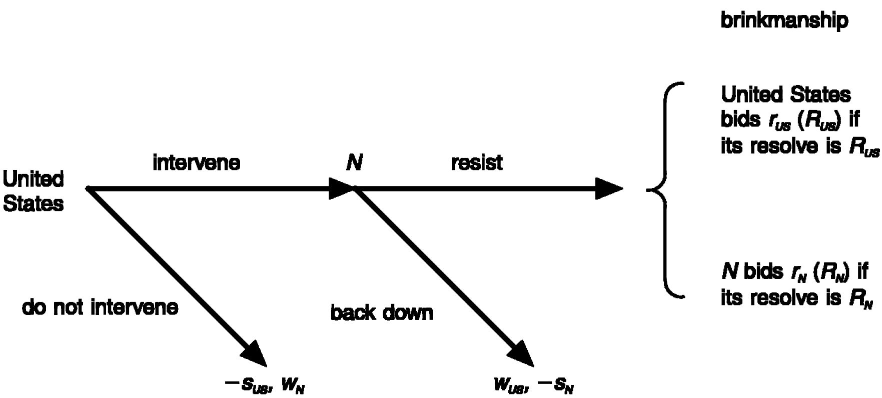
One conclusion follows immediately. As noted above, there are no crises when the balance of resolve is clear. Knowing that it cannot prevail, it is dangerous and pointless for the less resolute state to run any risk, and it acquiesces to the more resolute state. Moreover, in most cases in which the survival of a small state´s regime is at issue, the balance of resolve will clearly favor that regime. Its willingness to run risks to stay in power will generally dominate the United States´ willingness to run risks to depose it. Hence the spread of nuclear weapons will strongly inhibit any U.S. effort to use its military superiority to overthrow the regime of another nuclear state.[40]

The situation is more complex if the survival of the regime of the small nuclear state is not immediately at issue. In these circumstances, the balance of resolve may be unclear, and each state may be willing to press the crisis to coerce the other into backing down. How does the prospect that the United States may be facing a rogue capable of launching a nuclear attack on U.S. soil affect the dynamics of this interaction?[41] The brinkmanship model of a confrontation between the United States and a small nuclear state in Figure 4 addresses this question.

In this game, the United States has to decide whether to intervene against a regional nuclear state, and that state has to decide whether to resist if the United States intervenes. The states engage in brinkmanship only if neither backs down. Modeling these two decisions explicitly makes it possible to see

38. Sagan and Waltz, *The Spread of Nuclear Weapons,* pp. 20–21.
39. Ibid., pp. 23–24.
40. Reaching a similar conclusion are Betts, "Universal Deterrence or Conceptual Collapse," pp. 65–66; and Waltz in Sagan and Waltz, *The Spread of Nuclear Weapons,* pp. 28–29.
41. Barry R. Posen, "U.S. Nuclear Policy in a Nuclear-Armed World, Or What If Iraq Had Had Nuclear Weapons?" in Utgoff, *The Coming Crisis,* pp. 157–190, considers the case in which a nuclear-armed adversary has the capability to deliver a few nuclear weapons regionally.

Figure 4. A Brinkmanship Model of a Nuclear Confrontation.



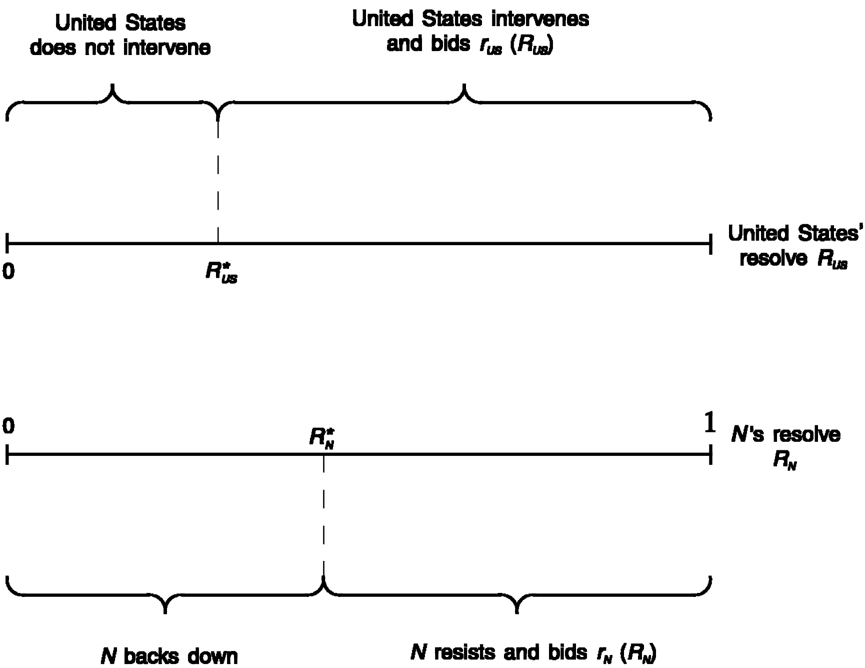
how rogue states and NMD affect the likelihood that the United States intervenes and that its adversary resists.

As for the payoffs of the game, the United States pays the cost of submitting, s_{US}, and the small nuclear state receives the payoff to prevailing, w_N, if the United States decides not to intervene. If the United States decides to oppose the other state, that state has to decide whether to back down or resist. The former ends the confrontation with the United States' receiving w_{US}, for prevailing and the other state's paying s_N for submitting. If the small nuclear state resists, the states engage in brinkmanship, and the state that is willing to tolerate the greatest risk prevails (assuming that events do not go out of control). If both states are willing to hazard the same level of risk, the status quo on the ground remains in place. The payoffs to these outcomes are defined as they were in the discussion of brinkmanship above with the additional fact that because the fighting that generates the risk is costly, each state pays a cost if the states find themselves engaged in brinkmanship. (See the appendix for a formal specification of these payoffs.)

Figure 5 illustrates the equilibrium of the game. The United States acquiesces if its resolve is below a critical level R^*_{US}. If the United States' resolve is above this threshold, it intervenes against its nuclear adversary and bids r_{US} if the other state does not back down. As in the brinkmanship game analyzed above, the risk that the United States is willing to run depends on its resolve, R_{US}; its uncertainty about the other state's resolve, τ_N; and that state's uncertainty about the United States, τ_{US}.

The small nuclear state pursues a similar strategy. It backs down in the face

Figure 5. The Equilibrium Strategies.



of U.S. opposition if its resolve is below a critical level R^*_N. If its resolve is above this level, it resists the United States and bids r_N in the ensuing brinkmanship conflict. This risk depends on its resolve, its uncertainty about the United States´ resolve, and the uncertainty surrounding its own resolve.

The appendix derives explicit expressions for the states´ bidding strategies and their thresholds. These expressions make it possible to trace the effects of changes in the states´ levels of resolve or their uncertainty on (1) the probability that the United States is deterred from opposing a nuclear-armed adversary, (2) the likelihood that an adversary backs down if the United States does intervene, and (3) the chances that the crisis goes out of control if neither state backs down. These expressions also provide an explicit way of assessing what happens if the small nuclear state is a rogue in U.S. eyes.

Rogue or backlash states may be defined conceptually as states that are more determined to prevail and therefore are more resolute than ordinary states.[42]

42. The present analysis focuses on the effects that rogues have on the dynamics of brinkmanship and not on whether any particular state actually is a rogue.

Because they are more determined to prevail, rogues are believed to be "especially willing to take risks."[43] That is, the leaders of rogue states may be willing to risk retaliation "in circumstances where more traditional, or at least more cautious leaders would not."[44] To formalize this idea, recall that the parameter τ_N represents the United States´ beliefs about the overall level of the other state´s resolve. The higher τ_N, the more resolute the United States believes the other state to be. Consequently, we can trace the effects of facing a more rogue-like adversary by seeing what happens as τ_N increases.

Intuitively, we might expect that the tougher a regional nuclear adversary is believed to be, the less likely the United States is to intervene and the more likely that adversary is to resist. Figure 6 tracks these expectations. The United States opposes the regional power if its resolve is above the threshold R^*_{US}. Thus, the likelihood that the United States intervenes is the probability that its resolve is above this threshold. This threshold increases and the probability of U.S. intervention falls as the regional nuclear power becomes tougher (i.e., as τ_N increases). Conversely, the threshold R^*_N falls and the probability that the adversary would resist U.S. intervention rises as that adversary becomes more of a rogue.

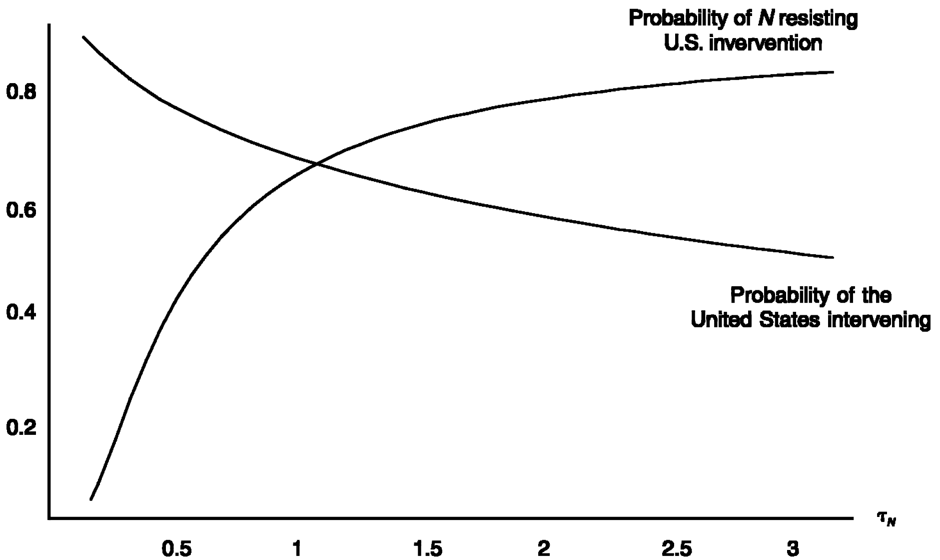
Figure 7 traces effects that a rogue has on stability. The chances that a confrontation eventually spirals out of control are very small when an adversary is likely to be irresolute (τ_N is small). In these circumstances, the balance of resolve, although not completely clear, is very likely to favor the United States and the risk is low. As an adversary becomes "tougher" or more of a rogue (i.e., as τ_N rises), the balance of resolve begins to blur and the risk increases. But, as that state becomes still more resolute, the balance of resolve becomes clearer (and clearly in favor of the rogue) and the confrontation becomes more stable. The net result of these effects is that the more resolute or "roguish" an adversary is perceived to be, the better off it is and the worse off the United States is (i.e., the higher the rogue´s expected payoff and the lower the United States´ payoff).

Brinkmanship and NMD

How does NMD affect the dynamics of brinkmanship? As is shown below, NMD effectively makes the United States more resolute. Consequently, the

43. Walt, "Containing Rogues and Renegades," p. 193.
44. Walter A. Slocombe, "The Administration´s Approach," *Washington Quarterly,* Vol. 23, No. 3 (Summer 2000), p. 80.

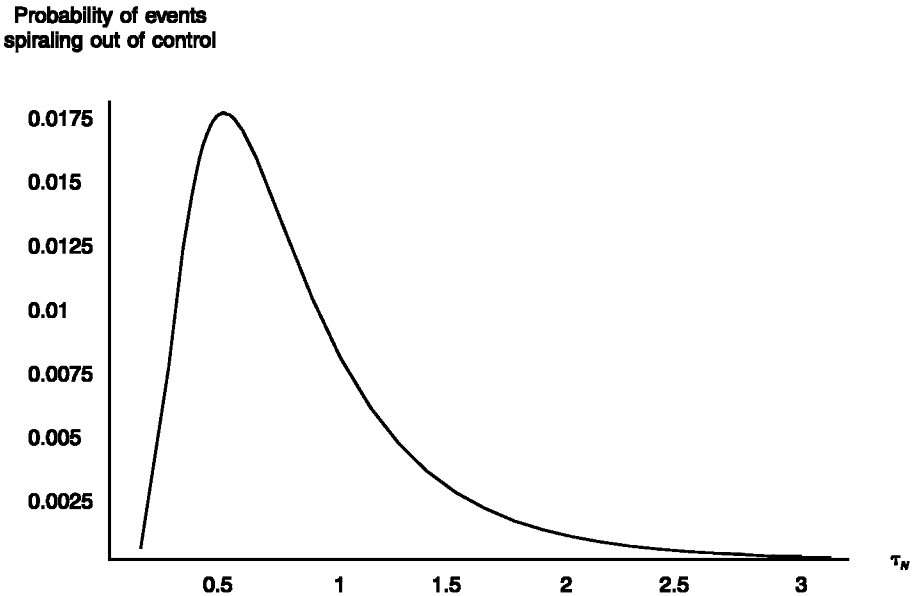
Figure 6. S_2's Perceived Resolve and the Probabilities of Intervention and Resistance.



United States becomes more likely to oppose a nuclear adversary and more willing to tolerate a higher risk (i.e., to bid more) if the other state does not back down. In response, the other state is more likely to back down and less willing to hazard the risk of escalation. The influence of NMD on stability depend on its effectiveness. As this effectiveness increases, the United States becomes more willing to run greater risks, and this begins to blur the balance of resolve between the United States and a rogue. This greater uncertainty surrounding the balance of resolve increases the chances of events going out of control and of a nuclear attack against the United States. But as the effectiveness continues to rise, the balance of resolve becomes clearer and the risk of nuclear attack begins to decline. Of course, the probability that a nuclear attack will actually reach the United States depends on both the likelihood of an attack and the effectiveness of NMD. The present analysis indicates that the probability that nuclear weapons will actually strike the United States in a confrontation with a rogue increases as NMD becomes more effective, until peaking at a very high level of effectiveness and then declining as NMD becomes still more effective.

The first step in tracing the consequences of NMD is to incorporate missile defenses into the model. To this end, note that the United States must still rely on deterrence to prevent an attack whether or not it has deployed NMD. Mis-

Figure 7. The Effect of S_2's Perceived Resolve on Stability.



sile defenses do not affect another state´s ability to launch an attack and, therefore, do not alter the basic process that generates the probability that events will go out of control that underlies brinkmanship. Rather NMD mitigates the consequences if events do go out of control and an adversary launches a nuclear attack. In particular, NMD reduces the cost to the United States in the event of such an attack.

This observation provides a point of departure for incorporating NMD into brinkmanship. Absent NMD, the cost to the United States if events go out of control is d_{US}. If the United States deploys NMD, the more effective it is, the lower this cost. Let e denote this effectiveness, where e can range from zero for a completely ineffective system to one for a perfect system.[45] Then the United States´ expected cost if attacked is $(1 - e)d_{US}$.

By reducing the costs if events go out of control, NMD increases the effective resolve of the United States. That is, missile defenses increase the maximum

45. How one actually assesses and characterizes the effectiveness of NMD is extremely complicated. A useful introduction is "Testing the NMD System: Requirements and Recommendations," in Andrew M. Slesser et al., *Countermeasures: A Technical Evaluation of the Planned U.S. National Missile Defense System* (Cambridge, Mass.: Union of Concerned Scientists, April 2002), pp. 91–102, http://www.ucsusa.org/publication.cfm?publicationID=348.

risk of attack that the United States would be willing to run in order to prevail. Indeed the better NMD works (i.e., the higher e), the more resolute the United States becomes.[46]

NMD, however, does not eliminate an adversary´s uncertainty about U.S. resolve. If an adversary were uncertain of the precise level of U.S. resolve absent NMD, it would still be uncertain of the precise level of U.S. resolve with NMD. NMD simply "shifts" the uncertainty that would exist about U.S. resolve absent NMD to higher levels. As a result, an adversary would believe that U.S. resolve was higher on average with NMD than it would have been without these defenses. If, as some argue, missile defenses enable the United States to retain its "freedom to respond to a regional crisis because they [the defenses] would negate the potential of regional aggressors with small, long-range missile forces to attack the U.S.,"[47] this is the way that NMD would do so.

Figures 8–10 illustrate the effects of NMD on crisis bargaining for differing levels of effectiveness e. For comparison, the reported design requirement for NMD is a "95 percent effectiveness with 95 percent confidence against a small-scale attack."[48] Former Secretary of Defense William Perry believes that "NMD could demonstrate on the test range a technical effectiveness of 80–90 percent" in a few years, but the operational effectiveness of the system will be much less.[49]

As NMD becomes more effective, the threshold of U.S. intervention, R^*_{US}, decreases and the United States becomes more likely to intervene. By contrast, the threshold of resistance R^*_N rises, and the regional nuclear adversary becomes less likely to resist intervention. Figure 8 traces these effects.

The figure also suggests that these effects are very small unless the defenses are very good (i.e., the probabilities of intervention and resistance stay relatively flat until the defenses become very good). If the defenses are no more than moderately effective, then the balance of resolve remains relatively clear and in the favor of the other state. Defenses that are only moderately good, therefore, do not make the United States significantly more likely to intervene,

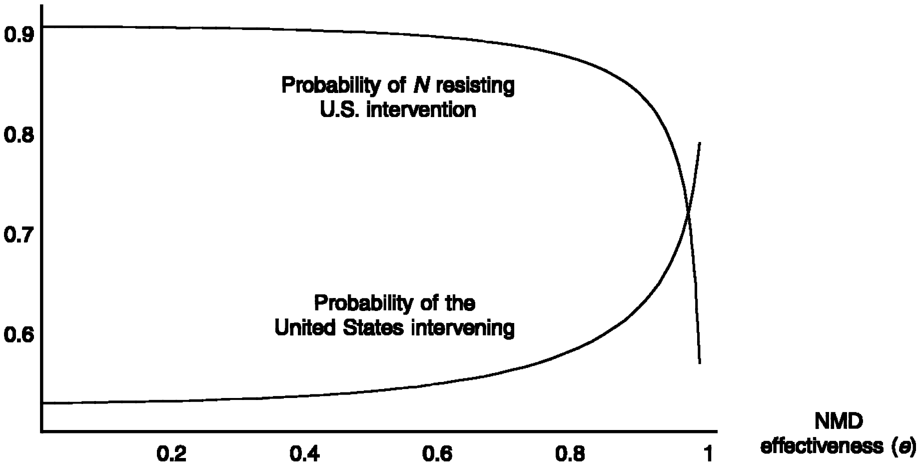
46. More precisely, the United States´ resolve with or without NMD remains $R_{us} = (w_{us} + s_{us})/(w_{us} + d_{us})$. The United States behaves, however, as if its resolve were $R_{us}(e) = (w_{us} + s_{us})/[w_{us} + (1 - e)d_{us}]$, if it has missile defenses with effectiveness e. Accordingly, $R_{us}(e)$ will be called the United States´ effective resolve. To ease the exposition, the adjective "effective" will be used only when needed to avoid ambiguity.

47. Slocombe, "The Administration´s Approach," p. 80.

48. Lewis, Gronlund, and Wright, "National Missile Defense," p. 126.

49. William J. Perry, "Preparing for the Next Attack," *Foreign Affairs,* Vol. 80, No. 6 (November–December 2001), pp. 64–80.

Figure 8. NMD Effectiveness and the Probabilities of Intervention and Resistance.

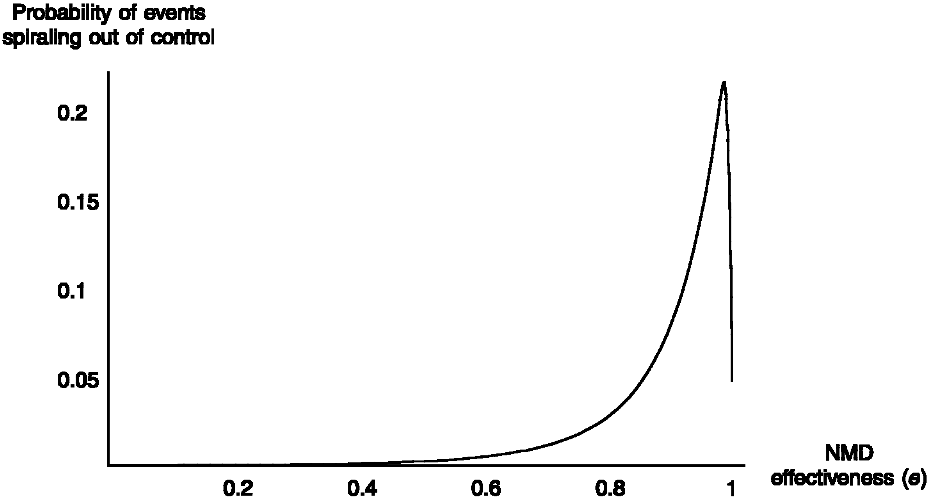


and they do not have much of an effect on an adversary's decision to resist if the United States does intervene.

Figures 9 and 10 highlight NMD's effects on stability. The former centers on the probability of events going out of control and of a nuclear attack on the United States. If the balance of resolve absent NMD clearly favors the regional nuclear state and NMD is ineffective, the United States will not push the crisis very hard and the risk of events spiraling out of control is small. As NMD becomes more effective, the United States' effective resolve increases, the balance of resolve begins to blur, and the probability of an attack rises. This probability continues to rise until NMD becomes very effective and then begins to fall as the balance of resolve begins to become clearer. NMD has to be virtually flawless before the probability of an attack drops below what it would have been without NMD.

The point at which the probability peaks depends on how resolute the small nuclear state is believed to be (i.e., on τ_N). The more resolute an adversary is believed to be, the more effective NMD must be before its increasing effectiveness stops blurring the balance of resolve and begins to clarify it. Indeed if the adversary is a rogue—as proponents of NMD often argue is likely to be the case—then missile defenses must be extremely effective before they begin to clarify rather than blur the balance of resolve. This suggests that NMD is likely to raise the risk of a nuclear attack on the United States in a confrontation with a rogue.

Figure 9. NMD Effectiveness and the Probability That a Confrontation Goes Out of Control.

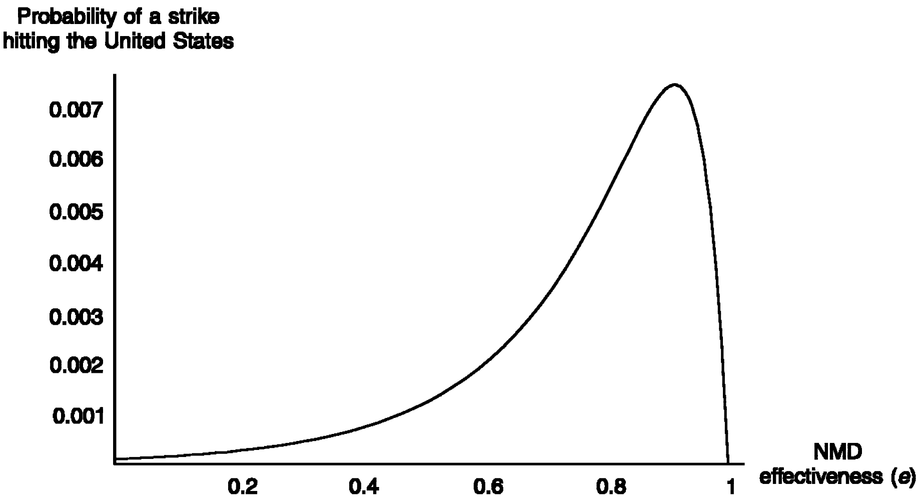


Of course, the probability that an attack penetrates U.S. missile defenses and reaches the United States depends both on the probability of an attack and on how well those defenses work. Figure 10 traces the effects of NMD on the probability that the United States will be struck by a nuclear weapon.[50] Once again, this probability initially rises and then eventually falls as NMD becomes more effective.

Figures 9 and 10 highlight an important and underappreciated or, at least, underemphasized policy trade-off. To the extent that missile defenses help to create more freedom of action for the United States and help it to achieve its ends in a confrontation, they do so by increasing the United States' effective resolve. As a result, the United States is willing to push the crisis harder: It is more likely to intervene, and it is willing to run a higher risk of events spiraling out of control. But there is a price to be paid. As missile defenses increase U.S. resolve, the balance of resolve blurs, and the probability of a nuclear strike on the United States rises. This increase is not the result of misperception or an overestimate of the effectiveness of the defenses.[51] It is the direct consequence

50. This probability is taken to be the chance that NMD fails, $1 - e$, multiplied by the probability of an attack.
51. For this argument, see Glaser and Fetter, "National Missile Defense and the Future of U.S. Nuclear Weapons Foreign Policy," p. 69.

Figure 10. NMD Effectiveness and the Probability of a Strike on the United States.



of pushing the crisis harder. On balance, NMD does increase the expected payoff to the United States, but at the cost of a higher probability of nuclear weapons striking the United States.[52]

Conclusion

Brinkmanship provides a model for the way that severe conflicts of interest between nuclear-armed states play out. Applied carefully and appropriately, it helps to explain the dynamics of Cold War crises and to clarify the policy trade-offs posed by the spread of nuclear weapons after the Cold War. In situations in which the balance of resolve is very clearly in favor of a small nuclear state, brinkmanship indicates that the small state will be able to deter the United States. Consequently, the spread of nuclear weapons is likely to give the regime of a small nuclear state—whether a rogue or not—the ability to deter the United States from trying to overthrow that regime. This moreover is likely to be the case even if the United States deploys an NMD, unless that system is virtually flawless. In situations in which the balance of resolve is more ambiguous, NMD increases the effective resolve of the United States. Defenses

52. Recall that the expected payoff is the payoff to prevailing, w_{us}, the cost of submitting, s_{us}, and the cost to suffering an attack, $(1 - e)\, d_{us}$, weighted by the probabilities of these outcomes.

that are very good would give the United States somewhat more freedom of action and make a rogue more likely to back down in a crisis. But there is a trade-off. The United States is more likely to achieve its ends because it is willing to press the crisis harder. This, however, is likely to raise the probability of a nuclear attack on the United States. Unfortunately, the relative stability between the United States and the Soviet Union during the second half of the Cold War provides a poor guide to the stability of a crisis between the United States and a nuclear-armed, regional adversary. Stability depends on what the future conflicts of interest turn out to be and how clear the balance of resolve is in those conflicts.

Appendix

This appendix formally describes an equilibrium of the brinkmanship crisis game illustrated in Figure 4 in which the players have to decide whether to back down before bidding. This game is more complicated than the one discussed earlier in the section "Nuclear Deterrence Theory Revisited" in which the players only bid, and the equilibrium of this simpler game follows immediately from the analysis of the more complicated game. The appendix concludes with a discussion of the parameter values used in the various figures and a table summarizing those values.

A FORMALIZATION OF BRINKMANSHIP

The brinkmanship crisis game explicitly models the United States' decision about intervening and the regional nuclear power's decision about resisting U.S. intervention. The states engage in brinkmanship only if the United States decides to intervene and the regional nuclear state decides to resist.

If the United States chooses not to intervene, it obtains $-s_{US}$ and the regional nuclear power receives w_N. If the United States does intervene and the other state backs down, the states receive w_{US} and $-s_N$, respectively. If neither backs down, the states' payoffs depend on their bids.

If the small nuclear state bids more ($r_N > r_{US}$), it prevails but has to run the risk set by the United States' bid. This yields an expected payoff of $(w_N - k_N)(1 - r_{US}) - r_{US} d_N$, where k_N is the cost of the conventional conflict that generates the risk. (This cost is included in d_N, which represents the total cost wrought by the United States in retaliation for a nuclear attack.) The United States' payoff if it is outbid depends on the effectiveness of its missile defenses.

Let d_{US} be its cost if events go out of control and it has no defenses. Then, its cost is $(1-e)d_{US}$, if it has defenses with effectiveness *e*. Given this cost, the United States´ expected payoff if it is outbid is $-(s_{US} + k_{US})(1 - r_{US}) - r_{US}\,[(1 - e)d_{US} + k_{US}]$, where k_{US} is the cost of the conventional conflict creating the risk.[53] If the United States bids more than its adversary ($r_{US} > r_N$), the states´ payoffs are $(w_{US} - k_{US})(1 - r_N) - r_N[(1 - e)d_{US} + k_{US}]$ and $-(s_N + k_N)(1 - r_N) - d_N r_N$. Finally, if both states bid the same risk, the situation on the ground prevails. The regional nuclear power receives the benefits of the (unmodeled) action that started the confrontation and forced the United States to decide whether or not to intervene. Formally, the United States and the regional power obtain $-(s_{US} + k_{US})(1 - r_{US}) - r_{US}[(1 - e)d_{US} + k_{US}]$ and $(w_N - k_N)(1 - r_{US}) - d_N r_{US}$.

The states are also unsure of each other´s resolve. More precisely, each state is unsure of the other´s payoff to prevailing, which in turn induces uncertainty over resolve. To wit, the United States believes that $w_N \geq 0$ is distributed according to the Weibull distribution $1 - \exp\left[-(1/\tau_N)\sqrt{(w_N + s_N)/(s_N + d_N)}\right]$ (where s_N will be normalized to zero). This implies that *N*´s resolve, which is defined to be $R_N = (w_N + s_N)/(w_N + d_N)$, is distributed according to $G_N(R_N) = 1 - \exp\left[-(1/\tau_N)\sqrt{R_N/(1 - R_N)}\right]$. Similarly, the regional nuclear state believes that $w_{US} \geq 0$ follows the Weibull $1 - \exp[-(1/\tau_{US})\sqrt{(w_{US} + s_{US})/(s_{US} + d_{US})}\,]$ and, consequently, U.S. resolve R_{US} is distributed according to $G_N(R_N) = 1 - \exp - \left[(1/\tau_N)\sqrt{R_N/(1 - Rs_N)}\right]$.[54]

An equilibrium is a pair of strategies such that each player´s strategy is optimal against the other´s. In the present game, a pair of strategies is a combination of two thresholds, R^*_{US} and R^*_{US}, and two functions, $r_{US}(R_{US})$ and $r_N(R_N)$, which describe when a state intervenes or resists and how much it bids if it does intervene or resist. In particular, the United States intervenes only if its resolve is at least as large as R^*_{US}, in which case it bids $r_{US}(R_{US})$. The regional nuclear state resists and subsequently bids $r_N(R_N)$ if $R_N \geq R^*_N$ and backs down otherwise. These strategies constitute a perfect Bayesian equilibrium of the

53. The costs s_{US} and s_N are normalized to be zero. Nevertheless, it will be often be clearer to include these costs explicitly in the analysis.

54. The game is well defined for any distribution of $w_N \geq 0$ and $w_{US} \geq 0$, but the states´ bids $r_N(R_N)$ and $r_{US}(R_{US})$ can be characterized explicitly only in special cases such as the Weibull distribution. See Merran Evans, Nicolas Hastings, and Brian Peacock, *Statistical Distributions*, 3d ed. (New York: Wiley, 2000), for a discussion of the properties of this distribution.

game if neither state, regardless of its resolve, has any incentive to deviate from its strategy given that the other player is following its strategy. This in turn implies that the strategies must satisfy four conditions.

In order to specify these conditions, it is helpful to specify R_{US}´s payoff to bidding *r.* (It is convenient to use "R_{US}" to refer to "the United States with a level of resolve equal to R_{US}.") This payoff is:

$$V_{US}(r \mid R_{US}) = -\left[(s_{US} + k_{US})(1-r) + r((1-e)d_{US} + k_{US})\right]\frac{1 - G_N(r_N^{-1}(r))}{1 - G_N(R_N^*)}$$

$$+ \int_0^{r_N^{-1}(r)} \left[(w_{US} - k_{US})(1 - r_N(R)) - r_N(R)((1-e)d_{US} + k_{US})\right]\frac{g_N(R)}{1 - G_N(R_N^*)}\, dR.$$

The factor in brackets in the first term is the United States´ payoff if *N* bids more than *r,* which in turn is weighted by the probability that *N* actually does bid more than *r.* This probability reflects the fact that the United States´ initial beliefs about its adversary´s resolve have been updated in light of the fact that this adversary has resisted U.S. intervention and thereby signaled that its resolve is at least R_N^*. Using Bayes´s rule, these updated beliefs are $[1 - G_N(R)]/[1 - G_N(R_N^*)]$. The expression in the brackets in the second term is the United States´ payoff if R_N bids $r_N(R_N) < r$, and the integral of this expression is the United States´ expected payoff given that *N* actually does bid less than *r.* Let $V_{US}^*(R_{US})$ be R_{US}´s maximum payoff—that is, what R_{US} obtains if it maximizes $V_{US}(r \mid R_{US})$ by making its optimal bid. The regional nuclear power´s payoffs $V_N(r \mid R_N)$ and $V_N^*(R_N)$ are defined in a parallel way.

Turning to the four equilibrium conditions, the first is that if the United States´ resolve is below R_{US}^*, then not intervening must maximize its payoff. That is, $-s_{US} \geq V_{US}^*(R_{US})$ for $R_{US} \leq R_{US}^*$. Second, backing down must maximize R_N´s payoff whenever $R_N \leq R_N^*$. Third, bidding $r_{US}(R_{US})$ must maximize R_{US}´s payoff whenever $R_{US} \geq R_{US}^*$: $V_{US}(r_{US}(R_{US}) \mid R_{US}) = V_{US}^*(R_{US})$. Finally, R_N´s bidding $r_N(R_N)$ must maximize its payoff: $V_N(r_N(R_N) \mid R_N) = V_N^*(R_N)$.

The first condition helps to characterize the threshold R_{US}^*. R_{US}^* must be indifferent between intervening and not; otherwise some R_{US} close to R_{US}^* could benefit by deviating from the strategy of intervening only if $R_{US} \geq R_{US}^*$, which would contradict the definition of an equilibrium. R_{US}^*´s indifference implies that its payoff to not intervening, $-s_{US}$, must equal its payoff to intervening. The calculation of the latter is based on the fact that the least resolute type that intervenes must bid zero in any equilibrium in which both states bid more than zero with positive probability. (This is a standard result in all-pay auctions.) Thus R_{US}^{*}'s payoff to intervening and subsequently bidding zero is $w_{US}G_N(R_N^{*}) - (s_{US} + k_{US})[1 - G_N(R_N^{*})]$. The first term in this expression is R_{US}^{*}'s payoff if N does back down weighted by the probability that N backs down, and the second term is R_{US}^{*}'s cost to bidding zero in the event that N does not back down weighted by the probability that N does not back down. Setting this payoff to $-s_{US}$, dividing through by d_{US}, and using the normalization $s_{US} = 0$ gives

$$\frac{k_{US}}{d_{US}} = R_{US}^{*}\left[\exp\left(\frac{1}{\tau_N}\sqrt{\frac{R_N^{*}}{1 - R_N^{*}}}\right) - 1\right]. \tag{A1}$$

The second condition helps determine R_N^{*}. As with R_{US}^{*}, R_N^{*} must be indifferent between backing down and resisting with a subsequent bid of zero. This implies $-s_N = (w_N^{*} - k_N)\beta - (1 - \beta)(s_N + k_N)$, where β is the probability that the United States bids zero if it meets resistance given that it has already intervened. Letting R_{US}^{**} be the highest level of resolve that bids zero, the probability β is given by $\beta = [G_{US}(R_{US}^{**}) - G_{US}(R_{US}^{*})]/[1 - G_{US}(R_{US}^{*})]$. Substituting the expression for β into the previous equation gives:

$$\frac{k_N}{d_N} = \frac{R_N^{*}}{1 - R_N^{*}}\left[1 - \exp\left(\frac{1}{\tau_N}\left(\sqrt{\frac{R_{US}}{1 - R_{US}^{*}}} - \sqrt{\frac{R_{US}^{**}}{1 - R_{US}^{**}}}\right)\right)\right]. \tag{A2}$$

The third and fourth equilibrium conditions provide a third equation that pins down R_{US}^{*}, R_{US}^{**}, and R_N^{*} as well as the functions r_{US} and r_N. Recall that bidding $r_{US}(R_{US})$ and $r_N(R_N)$ must respectively maximize R_{US}'s and R_N's payoffs. Consequently, $r_{US}(R_{US})$ must satisfy the first-order condition obtained by differentiating $V_{US}(r|R_{US})$ with respect to r and setting it equal to zero:

$$1 = \left(\frac{1}{1 - e}\right)\left(\frac{R_{US}}{1 - R_{US}}\right)\left(\frac{g_N(r_N^{-1}(r))}{1 - G_N(r_N^{-1}(r))}\right)\left(\frac{1 - r}{r_N'(r_N^{-1}(r))}\right).$$

Similarly, $r_N(R_N)$ must satisfy the first-order condition resulting from differentiating $V_N(r|R_N)$:

$$1 = \left(\frac{d_N}{d_N - k_N}\right)\left(\frac{R_N}{1 - R_N}\right)\left(\frac{g_{US}(r_{US}^{-1}(r))}{1 - G_{US}(r_{US}^{-1}(r))}\right)\left(\frac{1 - r}{r_{US}'(r_{US}^{-1}(r))}\right).$$

The solution to the system of differential equations defined by the first-order conditions and the restriction that R_{US}^{**} and R_N^{*} bid zero is:

$$r_{US}(R_{US}) = 1 - \exp\left[-\frac{\tau_{US}}{3\tau_N^2}\left(\frac{1}{1-e}\right)^2\left(1 - \frac{k_N}{d_N}\right)\left[\left(\frac{R_{US}}{1-R_{US}}\right)^{3/2} - \left(\frac{R_{US}^{**}}{1-R_{US}^{**}}\right)^{3/2}\right]\right]$$

$$r_N(R_N) = 1 - \exp\left[-\frac{\tau_{US}}{3\tau_{US}^2}(1-e)\left(1 - \frac{k_N}{d_N}\right)^{-2}\left[\left(\frac{R_N}{1-R_N}\right)^{3/2} - \left(\frac{R_N^{*}}{1-R_N^{*}}\right)^{3/2}\right]\right]$$

where

$$\frac{R_N^*}{1-R_N^*} = \left(\frac{\tau_{US}}{\tau_N}\right)^2\left(\frac{1}{1-e}\right)^2\left(1-\frac{k_N}{d_N}\right)^2\frac{R_{US}^{**}}{1-R_{US}^{**}}. \tag{A3}$$

Equations (A1)–(A3) can be solved numerically for the thresholds R_{US}^*, R_{US}^{**}, and R_N^* which along with the bidding strategies r_{US} and r_N constitute an equilibrium. (The functions r_{US} and r_N with e, k_N, k_{US}, R_{US}^{**}, and R_N^* set to zero define an equilibrium of the simpler game in which the states bid but do not first have to decide about intervening and backing down.)

PARAMETER VALUES

In order to plot the figures, some assumptions have to be made about the parameter values. The results reported are only intended to be suggestive. They are based on what seem to be reasonable assumptions, but other reasonable assumptions could have been made. The preceding formal discussion shows how to use the model with other parameter values.

Table 1 specifies the parameters in the figures. There are two types of parameters—cost parameters and belief parameters. The cost parameters k_{US}/d_{US} and k_N/d_N specify the cost of the regional conflict that generates the risk of disaster (k_{US} and k_N) relative to the cost incurred if events go out of control (d_{US} and d_N). These relative costs are taken to be 1 percent for the United States and 10 percent for the small nuclear state ($k_{US}/d_{US} = 0.01$ and $k_N/d_N = 0.1$). The idea here is that the costs of the regional conflict are much higher for the small nuclear state.

The belief parameters reflect the states´ uncertainty about each other. The figures are based on a "nominal" case in which the median of the distribution representing a state´s beliefs is 10 percent (i.e., a state believes that there is a 50-50 chance that the other´s resolve is 10 percent or less). This occurs at the value of τ that satisfies $0.5 = 1 - \exp\left[-(1/\tau\sqrt{R/(1-R)}\right]$ at $R = 0.1$, which implies $\tau \approx 0.48$. Figures 3 and 7 trace the effects of varying τ for one of the states

Table 1. Parameter Values Used in the Figures.

Figure	Parameter Values
Figure 2	$\tau_1 = \tau_2 \approx 0.48$
Figure 3	$R_1 = 0.1, \tau_1 \approx 0.14$
Figure 6	$\tau_{US} \approx 0.48, k_{US} / d_{US} = 0.01, k_N / d_N = 0.1$
Figure 7	$\tau_{US} \approx 0.48, k_{US} / d_{US} = 0.01, k_N / d_N = 0.$
Figure 8	$\tau_{US} \approx 0.48, \tau_N \approx 3.16, k_{US} / d_{US} = 0.01, k_N / d_N = 0.1$
Figure 9	$\tau_{US} \approx 0.48, \tau_N \approx 3.16, k_{US} / d_{US} = 0.01, k_N / d_N = 0.1$
Figure 10	$\tau_{US} \approx 0.48, \tau_N \approx 3.16, k_{US} / d_{US} = 0.01, k_N / d_N = 0.1$

over a range of 0.14 to 3.16. At $\tau \approx 0.14$, a state is 90 percent confident that its adversary´s resolve is 10 percent or less. At $\tau \approx 3.16$, a state is 90 percent confident that its adversary´s resolve is 10 percent or higher. Thus a state with a level of resolve of 10 percent is quite confident that it is more resolute than its adversary if $\tau \approx 0.14$, and it is also quite confident that its adversary´s resolve is greater than its resolve if $\tau \approx 3.16$.