



---

Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model

Author(s): James D. Fearon

Source: *The Journal of Conflict Resolution*, Vol. 38, No. 2, Arms, Alliances, and Cooperation: Formal Models and Empirical Tests (Jun., 1994), pp. 236-269

Published by: [Sage Publications, Inc.](#)

Stable URL: <http://www.jstor.org/stable/174295>

Accessed: 24/10/2013 18:56

---

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

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



*Sage Publications, Inc.* is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Conflict Resolution*.

<http://www.jstor.org>

# Signaling versus the Balance of Power and Interests

## AN EMPIRICAL TEST OF A CRISIS BARGAINING MODEL

JAMES D. FEARON

*University of Chicago*

---

---

Conventional wisdom holds that in international disputes, a state's military threats are more likely to work the more the state is favored by the balance of power or the balance of interests. Analysis of a game-theoretic model of crisis signaling substantially refines and revises this claim. Due to selection effects arising from strategic behavior, measures of the relative strength of a defender's interests that are available before a crisis begins (*ex ante*) should be related to the *failure* of the defender's threats during the crisis. *Ex ante* measures of the defender's relative military strength should correlate with the success of the defender's crisis threats, but due to strategic dynamics that are not grasped by the standard arguments. A reanalysis of Huth and Russett's data on immediate deterrent threats lends support for these and other hypotheses drawn from the game-theoretic treatment.

---

---

Among the most important questions in social science are the causes and effects of threats and force. In international politics we are particularly interested in when threats protect the state and when, by contrast, they set off a spiral of counterthreats that leave both sides worse off than they would have been had the state adopted an alternative policy.

—Robert Jervis (1989, 183)

Jervis's question about the effects of threats in international politics seems ripe for theoretically informed empirical analysis. On the one hand, we have a growing theoretical literature that uses recent developments in game theory

---

AUTHOR'S NOTE: This article is based on chapter 5 of J. Fearon, *Threats to Use Force: The Role of Costly Signals in International Crises*, Ph.D. Dissertation, University of California, Berkeley, 1992. For helpful comments on the article I wish to thank Paul Huth, Bruce Russett, Barry O'Neill, and conference participants. I am especially indebted to Paul Huth and Bruce Russett for access to their data on immediate deterrence.

JOURNAL OF CONFLICT RESOLUTION, Vol. 38 No. 2, June 1994 236-269  
© 1994 Sage Publications, Inc.

236

to study the evolution and dynamics of international disputes (e.g., Nalebuff 1986; Powell 1990; Morrow 1989b; Fearon 1990, forthcoming-a; Wagner 1991; Kilgour 1991; Bueno de Mesquita and Lalman 1992). On the other hand, we have seen slow but steady improvements in the number and quality of data sets on international confrontations (e.g., Gochman and Maoz 1984; Brecher, Wilkenfeld, and Moser 1988; Huth and Russett 1988; Leng 1993). But with the notable exception of Bueno de Mesquita and Lalman (1992), there have been few efforts to test theoretical results empirically using these data sets. Relevant to the main concerns of this article, there have been virtually no efforts to draw specific hypotheses about the effects of threats from the new game-theoretic crisis models *and* to evaluate them empirically.

There are at least two reasons for this gap. First, as yet no crisis data set has been constructed with any game-theoretic model (or models) in mind. In consequence, distinctions that appear critical from the theoretical perspective provided by various models are not recognized in organizing crisis data, and evidence that might help evaluate them is not collected. This makes assessing theoretical results difficult by making it more difficult to translate easily between the theory and the data.

Second, modelers have probably not done enough to draw out hypotheses that can be tested using simple, plausible, and readily available measures. For example, given the importance of “power” and “interest” variables in informal arguments about international politics, it would seem desirable for a model to yield results on how observable and measurable aspects of military power and foreign policy interests would affect leaders’ choices in disputes. But this can be a challenge for a game-theoretic analysis, in which *unobservable* factors such as private information about capabilities or “resolve” often figure prominently.

In this article, I use an incomplete-information model of an international crisis to develop hypotheses on the impact of observable dimensions of military power and national interest on threat making in crises. The hypotheses are evaluated using Huth and Russett’s (1988) data on a particular class of international disputes—58 cases of “extended immediate deterrence” since 1885.

More than any other I know of, Huth and Russett’s (1988) crisis data are structured in a way appropriate for evaluating a limited-information game model of crisis bargaining. In the first place, Huth and Russett code crises as sequential, step-by-step events. In each of their cases, a “challenger” issued a threat to a “protégé” state, and a “defender” replied by issuing a deterrent threat on behalf of the protégé. Huth and Russett ask about the next steps in the sequence—whether the challenger decided to act on its initial threat, and whether the defender responded with military force if the challenger acted.

These coding rules represent crises in essentially the same way that the model summarized here does.<sup>1</sup> Second, Huth and Russett have measures for the relative military capabilities of the challenger, defender, and protégé, and on the extent of the defender's interest in the protégé.

Finally, the data are well suited for asking about the effects of threats made in international crises and for comparing hypotheses derived from the game-theoretic approach to older and more mainstream arguments. Huth and Russett (1988) and Huth (1988b) used the data to test a number of hypotheses on the conditions under which a defender's "immediate deterrent" threat is most likely to succeed or fail—that is, what predicts whether the challenging state acted or did not act on its initial threat against the protégé? Most of their hypotheses are based on informal rationalist arguments about the impact of the balance of capabilities and the balance of interests on the credibility of a threat. The central idea is that the defender's deterrent threat should be more likely to succeed the more the defender is favored by the balance of capabilities or interests. This idea is both highly intuitive and widely accepted by students of international politics (Fearon 1992, chap. 2).

The game-theoretic analysis sketched below substantially revises and qualifies these claims about the effects of relative power and relative interests in crisis bargaining. The main problem with the standard arguments is that they do not grasp the consequences of strategic dynamics linked to the fact that crises are sequences of decisions to threaten or escalate. In particular, the full empirical implications of strategic behavior by challenging states is missed. To the extent that relative capabilities and interests are observable before a crisis begins, rational challengers should take these into account. When the observable balance of interests favors the defender, only relatively resolved challengers will choose to threaten, implying that the defender's effort at immediate deterrence will be relatively unlikely to succeed (contrary to the standard hypothesis). When the observable balance of capabilities favors the defender, challenges will tend to occur on issues that are of initially doubtful interest to the defender. Hence a strong deterrent signal by the defender will be relatively likely to work in response, but due to the challenger's initial beliefs and choice of issue rather than (directly) due to the defender's superior military power.

Taken to the data, these hypotheses fare quite well (although there is one variable that acts anomalously) and yield an interpretation of the evidence quite different from that based on the standard balance of capabilities and

1. For more extended theoretical treatments of the model employed in this article, see Fearon (1990, 1992); for a more general formulation, see Fearon (forthcoming-a).

interest hypotheses. In brief, relative power and relative interests influence the effects of threats made in international confrontations, but in ways different and more subtle than the mainstream arguments envision.

The article proceeds in four sections. In the first, I briefly introduce the crisis bargaining model behind the empirical analysis. The second section considers some testable hypotheses based on equilibrium analysis of the model. In the third section, I reanalyze and reinterpret Huth and Russett's (1988) data on immediate deterrence in light of the game-theoretic arguments and hypotheses. A fourth section concludes.

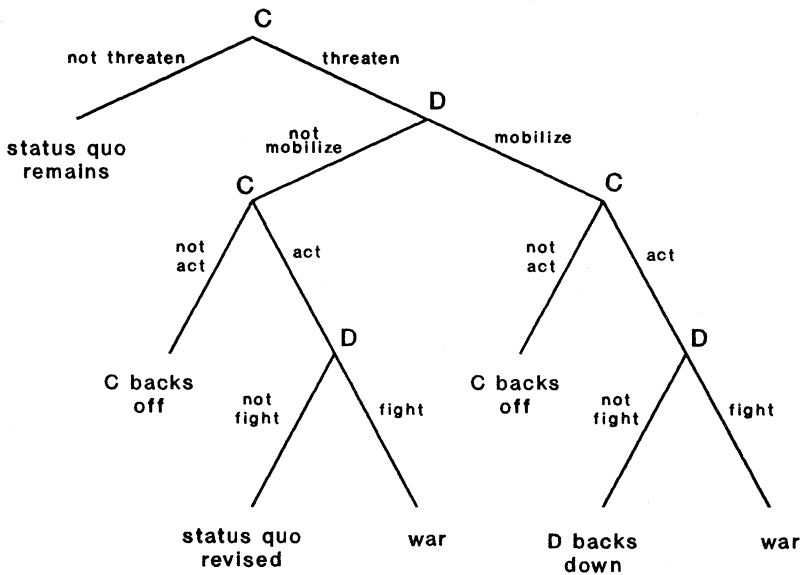
### A SIGNALING MODEL OF CRISIS BARGAINING

Since George and Smoke (1974), empirical work on international crises has emphasized that they are sequential in nature—one state challenges, the other responds, the first replies again, and so on until either war, capitulation, or some settlement is reached. This section presents and analyzes a model of a crisis as a four-step sequence. A challenger and a defender alternate in making threats or warnings; each begins the crisis uncertain about the adversary's willingness to fight over the issues at stake. Because the focus of the article is empirical and to save space, I do not fully characterize equilibrium results for the game (see Fearon 1990, 1992, chap. 4).

*The structure of the crisis.* The model depicts crises as having four principal stages or steps, as illustrated in Figure 1. In the first step, the challenger considers whether to take an action that explicitly or implicitly threatens another state. If the challenger decides not to threaten, the status quo prevails. If threatened, the defender chooses in the second step whether to respond with some threat or warning of its own. The response might be an action such as mobilization, a public declaration of intent to resist if the challenger chooses to act on its threat, or a declaration of intent to restore the status quo ante if the challenger has already begun changing the status quo.

In the third step, after observing the defender's response, the challenger chooses whether or not to act on its initial threat, or to continue if it had already begun. If the challenger chooses not to act, the crisis ends. Finally, in the fourth step, if the challenger acts, the defender chooses whether or not to resist with military force.

The model is much simpler than any particular historical case of crisis bargaining. In the real world, states have opportunities to send a greater variety of messages to each other through multiple channels; the choices of



**Figure 1: A Signaling Model of Crisis Bargaining**

more than two states may be relevant; and policies are produced and carried out not by unitary actors, but by organizations. Nonetheless, informal versions of this model appear repeatedly in empirical studies of crises, including George and Smoke (1974, 101-3), Huth and Russett (1984, 1988, 1993), Huth (1988b, 20-25), and Lebow and Stein (1990b). It also captures, in a natural way, a conceptual distinction that has played a major role in recent empirical studies of crisis bargaining. Morgan (1977, chap. 2) distinguished between general deterrence—which holds if no threats are issued between states that are involved in a generally adversarial relationship—and immediate deterrence—which becomes an issue after one state has decided to threaten or to take other actions that suggest the possible use of force. This distinction presupposes a sequential, step-by-step pattern for crisis bargaining like that represented in Figure 1. There, general deterrence fails if the challenger decides to threaten rather than accept the status quo. Immediate deterrence fails if the challenger acts on its initial threat after observing a tough response by the defender, such as mobilization.

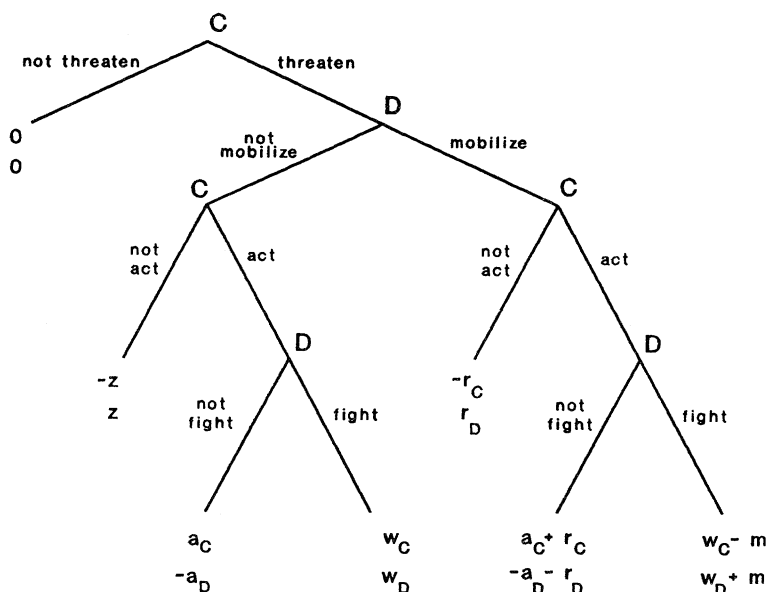


Figure 2: The Signaling Game with Payoffs

*Payoffs.* For convenience, status quo payoffs may be normalized to  $(0, 0)$ , where the first component is the challenger's and the second the defender's expected utility for living with the status quo. These, as well as the other payoffs, are depicted in Figure 2.

If war results when the defender has not mobilized, challenger and defender receive their expected utilities for conflict,  $(w_C, w_D)$ . If war occurs after military preparation by the defender, the challenger's value for conflict is reduced and the defender's increased by  $m \geq 0$ . Thus they receive  $(w_C - m, w_D + m)$ .

I have argued elsewhere that in most crises, threatening or making a "show of force" generates *audience costs* that a leader might suffer if he backed down later (Fearon 1990, 1992, forthcoming-a). These costs arise from the action of domestic political audiences concerned with whether the leadership is successful or unsuccessful at foreign policy. In the game, if the challenger backs down after the defender mobilizes, it will suffer an audience cost  $r_C > 0$ , whereas the defender enjoys a "foreign policy success" worth  $r_D > 0$ .

Payoffs are then  $(-r_C, r_D)$ . If the challenger backs down when no show of force has been made, it bears a small cost  $z$ , whereas the defender reaps a small benefit  $z$ , where  $0 - z < r_C$ . So  $(-z, z)$  results.<sup>2</sup>

Finally, if the defender backs down after it has mobilized and after action taken by the challenger, it loses the value of the issue at stake,  $a_D > 0$ , along with its audience cost  $r_D$ . Conversely, the challenger gains  $a_C > 0$ , its intrinsic value for whatever is at stake, plus  $r_C$ , the extra benefit of having successfully stood up to the defender's attempt at "immediate deterrence." Payoffs for this outcome are  $(a_C + r_C, -a_D - r_D)$ . On the other hand, if the defender backs down without having made a show of force, it suffers only the loss of the issue at stake, and the challenger receives only its value for the issue itself. Thus payoffs are  $(a_C, -a_D)$ .

To summarize, three key variables define the states' preferences over the several possible crisis outcomes: the states' values for the issues at stake in the dispute ( $a_C$  and  $a_D$ ); the audience costs created by public displays of force ( $r_C$  and  $r_D$ ); and the states' values for the war outcome ( $w_C$  and  $w_D$ ). Each of these values is defined relative to a player's value for the status quo, which is normalized to equal 0.

*Equilibrium with incomplete information about values for war.* When all of the above payoffs are common knowledge, the game has three possible subgame perfect outcomes depending on the states' values for war. Either the challenger accepts the status quo; the challenger threatens and the defender cedes the issue at stake without resistance; or war occurs after mobilization by the defender. So, with complete information, the defender's counterthreat fails whenever it is actually made—challengers can anticipate the defender's actions and so will not threaten expecting resistance unless they are themselves willing to go to war. The question, "When do deterrent threats work?" becomes interesting only when we bring incomplete information into the picture: if challengers are uncertain about the defender's true willingness to fight over the issues in question, then there is a potential role for the defender's counterthreat to "signal resolve" to a challenger, so dissuading further aggressive action.

To model uncertainty of this sort, we suppose that before the crisis begins, the states' expected utilities for war,  $w_C$  and  $w_D$ , are distributed uniformly on the intervals  $[-j, a_C]$  and  $[-k, 0]$ , respectively. At the outset, challenger and defender are informed of their own value for war,  $w_C$  and  $w_D$ , but know only the distribution of their adversary's value. Thus the distribution of  $w_D$  on

2. No results are affected if these are  $(-z_C, z_D)$ .



$[-k, 0]$  represents the challenger's initial belief, or uncertainty, about the defender's willingness to fight, whereas the distribution of  $w_c$  on  $[-j, a_c]$  represents the defender's initial belief about the challenger's value for war. The parameters  $j$  and  $k$  (with  $k$  taken to be greater than  $a_p$  and both as common knowledge) determine the nature of the states' initial beliefs: the larger  $j$  or  $k$ , the greater the initial belief that the challenger or defender probably has a *low* willingness to go to war over the issues at stake.<sup>3</sup>

So defined, the game proves to have a unique sequential equilibrium distribution on outcomes for any set of parameter values. Equilibrium is characterized by piecemeal or step-by-step separation of states according to their privately known values for conflict. Relatively tough challengers (i.e., high  $w_c$ ) choose to threaten and then act on their threat. Relatively tough defenders choose to mobilize against a threat and then fight back if the challenger acts. Challengers with lower values for conflict choose either not to threaten or to try a limited probe—to threaten and then back off if they encounter an immediate deterrent threat by the defender. Defenders with lower willingness to fight choose either not to offer resistance to a threat or to try immediate deterrence and then back down if it fails.

A crisis in the incomplete-information game thus appears as a succession of "costly signals" that allow states to learn about their adversary's true willingness to use military force.<sup>4</sup> On seeing mobilization by the defender, the challenger correctly increases its belief that the defender might in fact be willing to use force. The reason is that tough defenders are more likely to mobilize than weak defenders because tough defenders do not have to worry about paying audience costs for backing down—they prefer to go to war if it comes to this. However, although learning does take place, it is not necessarily complete. On seeing a threat, the defender will typically be unable to tell with certainty whether immediate deterrence will succeed or fail if tried. Likewise, on seeing an immediate deterrent threat by the defender, the challenger may not be able to judge whether the defender would actually be willing to use force on behalf of the protégé.

3. An alternative specification would have the states uncertain about the adversary's value for the issue ( $a_c$  and  $a_p$ ), while values for war were common knowledge. The same qualitative results hold, at least for the case of one-sided incomplete information (challenger's value for conflict is known). I use the more complicated case of two-sided incomplete information here to make possible challenger self-selection into crises by type ( $w_c$ ).

4. On costly signals, see Spence (1973) and Fearon (1992) for discussion of signaling in international crises. The costs that make signaling informative in this model differ somewhat from the "classical" costly signals of Spence. In the classical case, the act of sending the signal is itself costly, whereas here it is the act of backing down after it has been sent that is costly. The term "bridge burning signals" has been suggested, but with incomplete information the receiver may be unsure whether the bridge is really burned.

Even so, states in the model can form expectations about the likelihood that their threats or counterthreats will succeed, and in equilibrium these probabilistic estimates are accurate. We can use these equilibrium probability estimates from the model to derive hypotheses on when threats used in crises are more or less likely to succeed. In particular, equilibrium results allow us to ask how the probability of immediate deterrence success—that is, the probability that the challenger backs down after seeing a deterrent threat by the defender—varies with the several parameters of the model.

In the next section, I develop and interpret some of these results on the effects of crisis threats, contrasting them with the “mainstream rationalist theories” that dominate the literature.

### COMPARATIVE STATICS AND HYPOTHESES ON IMMEDIATE DETERRENCE

Mainstream rationalist theories characterize international crises as contests decided by a critical factor—the side with more military capabilities, more resolve, or stronger intrinsic interests is predicted to prevail. Although the logic is not always spelled out, these arguments are typically defended in rationalist terms. The core idea is that (1) having more of the relevant critical factor (capabilities, resolve, etc.) allows a state to make a more credible threat to escalate a crisis to military conflict, and that (2) the side not favored by the balance of the critical factor will realize its disadvantage and so is more likely to back down.<sup>5</sup>

In their work on immediate deterrence, Huth and Russett (1984, 1988) and Huth (1988b) base most of their hypotheses on these arguments. In operational terms, they predict that, first, the greater the ratio of the defender’s capabilities to the challenger’s, the more likely is an immediate deterrent threat to succeed. Second, they argue that the stronger the defender’s level of interest in the protégé, the greater the chance of immediate deterrence success. I will refer to the twin claims that (a) a favorable balance of capabilities, and (b) a favorable balance of interests improve the prospects for immediate deterrence success as the mainstream hypothesis.

5. Mainstream rationalist theories are really a bundle of arguments found repeatedly in the literature on international conflict, rather than a single coherent theory. For examples, see Osgood and Tucker (1967), Jervis (1971), Maxwell (1968), George and Smoke (1974), Snyder and Diesing (1977, chap. 3), Huth and Russett (1984, 1988), Huth (1988b), Shimshoni (1988), and Pape (1990). Mainstream rationalist accounts of international crises are discussed at length in Fearon (1992, chap. 2).

Note that neither this hypothesis nor the argument from which it derives predicts any difference between general and immediate deterrence concerning the effect of the balance of capabilities or interests. Nor does it suggest any distinction between indicators of capabilities or interests that are available before the crisis begins, and information that emerges after the initial threat, in the course of the crisis.

Equilibrium results for the model outlined above suggest an alternative specification. In particular, comparative statics analysis draws attention to the importance of states' prior beliefs about their adversaries' willingness to use force. If crises are characterized by private information and costly signaling, then states will "select themselves" into or out of crises according to these prior beliefs, and this fact will have implications for subsequent inferences and choices. One consequence is that rationalist hypotheses that are true for general deterrence may be exactly reversed for immediate deterrence. For related reasons, information about, say, the balance of interests, may have a different impact on the probability of immediate deterrence success depending on whether it is known *ex ante* (before the initial threat) or *ex post* (after the initial threat).

In brief, equilibrium results indicate that the challenger's prior expectation that the defender prefers fighting to conceding the issue ( $a_D/k$ , in the model) will significantly influence the probability of both general and immediate deterrence success, although in opposite directions. The more the challenger initially expects the defender to prefer war to conceding the issue (the larger  $a_D/k$ ), the more likely is general deterrence to succeed, other things equal. But if general deterrence does fail, immediate deterrence will then be less likely to succeed, despite the defender's initial credibility. By the same token, if the challenger initially expected that the defender would probably prefer concessions to war, then general deterrence will be less likely to succeed, but subsequent efforts at immediate deterrence will be more likely to work.

The reason is that rational challengers will select themselves into a crisis according to their beliefs about the defender's preference for war versus concessions, and will do so in a manner that influences the probability of immediate deterrence success (Fearon forthcoming-b). When the defender is initially expected to prefer war to backing down, only highly motivated, hard-to-deter challengers (high  $w_C$ ) will choose to threaten in the first place. When the defender is initially expected to probably prefer concessions to war, then the incentive for "opportunistic," probing challenges is increased; in this set of cases (low  $a_D/k$ ), the challengers will tend on average to have low values

for conflict and so will be relatively easily dissuaded by an immediate deterrent threat.<sup>6</sup>

One major implication is that empirical evaluations of immediate deterrence ought to specify whether information about relative interests or capabilities is available before or after the initial challenge. Consider a measure of the defender's interest in the protégé that is available *ex ante*—for example, the presence of an alliance between the two. We would expect such a measure to be positively related to the challenger's prior belief that the defender might be willing to use force on behalf of the protégé. Thus it should have a positive impact on the likelihood of general deterrence success, but a negative impact on immediate deterrence—challengers who threaten protégés with allies will, on average, be more highly motivated and less deterrable.<sup>7</sup> By contrast, information about the defender's commitment to the protégé that emerges after an initial threat obviously cannot influence the challenger's prior beliefs or general deterrence. It would, however, be expected to influence positively the likelihood that an immediate deterrent threat would work.<sup>8</sup> This argument is summarized in hypothesis 1.

Hypothesis 1: Measures of the strength of the defender's interest in the protégé that are available before a crisis begins should be related to general deterrence success, but to immediate deterrence failure. Measures of defender interest revealed during a crisis should be related to immediate deterrence success.

Understanding how, by this theory, a measure of the balance of capabilities will be related to immediate deterrence is more involved. It is natural to suppose that the more the *ex ante* balance of forces favors the defender, the greater will be the challenger's prior belief that the defender might be willing to use force, for a given issue  $a_D$ .<sup>9</sup> If this were the only effect of the balance

6. In the model, the equilibrium probability that the challenger will act on its initial threat (conditional on a threat followed by mobilization) weakly increases as  $a_D/k$  increases. See Fearon (1990, 1992, chap. 4) for details. An equivalent result emerges in a more general model (Fearon forthcoming-a) that allows for an arbitrary distribution of states' values for war: a rightward shift in the distribution of a state's possible values for war increases the chance that its opponent will not escalate the dispute, but it also increases the chance that the opponent will not back down if it does choose to escalate.

7. Huth and Russett (1984) reported a negative effect of alliance ties on immediate deterrence success and gave a brief "selection effect" explanation. See, also, Huth and Russett (1993).

8. In the model, for example, the information revealed when the defender mobilizes lowers the probability that the challenger will act on its threat. It can be shown that if the defender chooses not to mobilize in response to a threat, the challenger will always act and the defender will not resist (Fearon 1990, lemma 1).

9. In the model, increasing relative capabilities in favor of the defender means shifting the distribution of the defender's values for conflict ( $w_D$ ) upward while shifting the distribution of challenger values ( $w_C$ ) downward.

of forces on prior beliefs, then we would make the same prediction as for measures of the balance of interests: ex ante indicators of relative capabilities should be related to general deterrence success, but immediate deterrence failure, while ex post indicators should be related to immediate deterrence success.

However, prior expectations about the balance of forces will also influence the interests on which challengers will choose to threaten. For example, in the model the probability that a militarily tough defender will be threatened on an important issue (large  $a_D$ ) is lower than the probability that this defender would be challenged on an issue of lesser importance. Intuitively, a state not favored by the balance of capabilities will be reluctant to challenge a vital interest of the defender. Ex ante measures of the balance of capabilities will thus be related to the stakes of a dispute,  $a_C$  and  $a_D$ , in a way that affects the chances for immediate deterrence success. The more the ex ante balance favors the defender, the less likely a challenge on a vital interest; hence the lower the expectation that the defender will prefer war to concessions; hence a greater chance of immediate deterrence success. If the choice of issue by the challenger is endogenous, then across cases, observable measures of the defender's relative military strength will be related to the success of immediate deterrent threats.<sup>10</sup>

Thus the ultimate prediction is the same as that given by the mainstream hypothesis: the more the (ex ante observable) balance of forces favors the defender, the more likely is the defender's threat to work. However, the reasoning behind the prediction is quite different. In the mainstream theory the challenger belatedly recognizes that it has inferior military capabilities, and so backs down. In the signaling model, issues and strategic behavior, rather than relative capabilities per se, drive the result. The challenger is more likely to back down when not favored by the military balance because it is more likely to challenge when it is highly uncertain about the defender's willingness to respond, as summarized by hypothesis 2.

Hypothesis 2: The more the ex ante balance of capabilities favors the defender, the more challengers will threaten only on issues of dubious importance to the defender, and in consequence, the more likely is a costly signal (an immediate deterrent threat) to succeed. Ex post measures of the relative military strength

10. Ideally, the choice of issue by the challenger should be made endogenous in the model, rather than deducing the effects of relative power on issue selection from comparative statics. An appropriate modification of the model given above is to allow the challenger to select  $a = a_C = a_D$  in the first step of the crisis. In the case of one-sided incomplete information (the defender's value for conflict is privately known), the proposition given in the text holds: The more the observable balance of capabilities favors the defender, the smaller the initial demand  $a$  by the challenger, and the greater the probability of immediate deterrent success. Two-sided incomplete information introduces refinements-related complications that I consider in work in progress. For another crisis bargaining model that has the property that greater relative power by the defender reduces the equilibrium demands made of it, see Fearon (1992, chap. 1).

of the defender revealed during a crisis should also increase the chance of immediate deterrent success.

So far, I have focused on the distinction between *ex ante* and *ex post* measures of relative capabilities or interests, and how these should be expected to influence the outcomes of the defender's threats. *Ex ante* measures, I have argued, influence immediate deterrence via their influence on the challenging state's prior beliefs about the defender's resolve.

Another important influence on the challenger's prior beliefs is the history of relations between defender and challenger. If there was a previous crisis between the two, this implies that there is at least one issue between them on which both have high enough values for conflict to lead them to incur the costs and risks of crisis bargaining. In the equilibrium of the signaling game, states with the lowest values for conflict on an issue either do not challenge or do not respond with a costly signal if challenged. Thus a previous confrontation may reveal something about the states' private information: regardless of which side backed down, the mere fact that there was a resisted challenge means that the states involved have higher than expected values for conflict on some issue. It suggests, further, that any subsequent challenge will be made with a higher than usual initial belief that the defender would be willing to resist.<sup>11</sup>

So if we compare a set of crises between states that had no prior conflict to a set in which the states involved did have some prior conflict, we should expect immediate deterrence to be less likely to work in the second set. A first crisis reveals information about the states' values for conflict, and this implies that a second challenge will be made with an initial belief about the chance of resistance that is on average higher than in cases without a prior conflict. Hypothesis 3 follows.

Hypothesis 3: Immediate deterrent threats will be less likely to succeed in crises between adversaries who faced each other in a previous crisis.

### **A REASSESSMENT OF THE EMPIRICAL EVIDENCE ON IMMEDIATE DETERRENCE**

This section uses the hypotheses and arguments given above to reevaluate Huth and Russett's (1984, 1988) data on the effects of immediate deterrent

11. To the extent that states anticipate future crises with each other, the reasoning here really should be checked against the results of repeated versions of the game. The one repeated crisis model in the literature (Morrow 1989a) lends support for the argument given in the text.

threats in international crises. I consider, in turn, the impact of the balance of capabilities, the balance of interests, and the history of a conflict, beginning in each case with a review of Huth's (1988b) and Huth and Russett's (1984, 1988) empirical findings.

### THE BALANCE OF CAPABILITIES

Huth (1988b) tests four measures for the balance of capabilities. Three are expressed as a ratio of the defender's and protégé's forces to the challenger's forces: (1) the long-term balance, measuring the two sides' overall military and industrial capabilities; (2) the short-term balance, measuring relative capacity to call up and mobilize troops in the space of a few months; and (3) the immediate balance, measuring the ratio of forces present at the point of conflict immediately prior to the onset of hostilities or retreat by either side. The fourth measure is a dichotomous variable: whether or not the defending state possessed nuclear weapons.

Note that the immediate balance is the only *ex post* indicator among the four. The long- and short-term balances and the defender's nuclear status are all observable or known prior to an initial threat by the challenger.

Drawing on work by Mearsheimer (1983), Snyder (1984), Posen (1984), and Van Evera (1984), Huth (1988b) presents a relatively nuanced version of balance of capabilities theory. The core idea is the mainstream argument given above: "The probability of deterrence success increases as the balance of military forces between attacker and defender shifts to the advantage of the defender" (p. 41). Huth elaborates on this by linking specific capabilities to specific military strategies. Following Mearsheimer (1983), he suggests that challengers might adopt three possible approaches: a "limited aims" strategy; a rapid offensive attack, or *blitzkrieg*; or a war of attrition.

Huth hypothesizes that the long-term balance will influence the challenger's calculations concerning a possible war of attrition, whereas the immediate and short-term balances will shape its considerations concerning a *blitzkrieg* or a limited-aims strategy. Further, because the consensus is that militaries prefer the offensive and wish strongly to avoid wars of attrition, Huth (1988b) expects that "The immediate and short-term balance of military forces will have a greater impact on deterrence outcomes than will the long-term balance of forces" (p. 41).

For the final element of the military balance—the defender's nuclear status—Huth (1988b, 42) originally argued that nuclear weapons should not be expected to influence immediate deterrence outcomes because the threat to use them is incredible, at least for cases of extended deterrence. A "combination of salient political, military, and ethical questions about the

immediate and long-term consequences of nuclear use by the defender against a non-nuclear power raise serious doubts as to the credibility of such a decision" (1988a, 428). This is a mainstream rationalist argument: more credible threats are more likely to succeed.

In the 1988 data analysis, all of Huth's expectations about the balance of capability measures were born out. Immediate deterrence appeared significantly more likely to succeed the more the immediate and short-term balance favored the defender. By contrast, the effects of the long-term balance and nuclear weapons on immediate deterrence could not be reliably distinguished from zero. In a later article, Huth (1990) altered his model specification to include interactive effects and then found evidence that possession of nuclear weapons had a significant positive impact on immediate deterrence success. He explained the finding with the suggestion that, due to psychological biases, nuclear threats may be more credible to leaders than they rationally ought to be (pp. 272-73).

### **Reassessment**

The hypotheses and arguments drawn from the signaling model suggest a different interpretation of these results. Whereas Huth distinguishes among his measures on the basis of hypothesized military and political attributes of the capabilities, I have argued that the distinction between *ex ante* and *ex post* measures is at least as important. As in the mainstream hypothesis, both *ex ante* and *ex post* measures of the ratio of defender to challenger forces should be positively related to immediate deterrence success. But the proposed mechanism differs. An *ex post* measure such as the immediate balance—the local balance of forces after the defender has responded to the challenger's initial threat—reflects new information about the defender's willingness to resist and also how effectively it would be able to resist. By contrast, *ex ante* measures such as the long- and short-term balances and the defender's nuclear status will be related to immediate deterrence success because they will act as proxies for the challenger's initial beliefs about the defender. When any of these balances strongly favors the defender, threats will most often have been made with the initial belief that the defender was probably not willing to use force on the issue. Hence a costly signal in response will be comparatively likely to succeed.

Evidence for this latter interpretation can be seen by looking at how the short-term balance variable divides the cases in Huth and Russett's (1988) sample. This variable is constructed as a ratio of mobilization capacities, based on the states' "standing ground and air forces and first class of trained reserves" (Huth 1988a, 432). Table 1 lists the set of cases in which the ratio



TABLE 1  
Cases For Which the Ratio of Short-Term  
Mobilization Capacities Is above Average (in favor of the defender)

<i>Case</i>	<i>Year</i>	<i>Challenger</i>	<i>Protégé</i>	<i>Defender</i>	<i>Outcome</i>
2	1885-1886	Bulgaria	Serbia	Austria-Hungary	Success
3	1886	Greece	Turkey	Britain	Success
5	1897	Greece	Crete	Britain/Turkey	Success
9	1903-1904	Columbia	Panama	United States	Success
18	1913	Rumania	Bulgaria	Russia	Success
20	1913	Serbia	Albania	Austria-Hungary	Success
25	1921	Panama	Costa Rica	United States	Success
28	1935-1936	Japan	Mongolia	Soviet Union	Success
39	1950	United States	North Korea	China	Failure
41	1957	Turkey	Syria	Soviet Union	Success
42	1961	Iraq	Kuwait	Britain	Success
43	1961	North Vietnam	Laos	United States	Success
47	1964-1965	North Vietnam	S. Vietnam	United States	Failure
51	1970	Syria	Jordan	Israel	Success
52	1971	India	Kashmir (Pakistan)	China	Success
54	1975	Morocco	W. Sahara	Spain	Failure
55	1975	Guatemala	Belize	Britain	Success
56	1977	Guatemala	Belize	Britain	Success
58	1983	Libya	Chad	France	Success

is above average for the sample (thus in favor of the defender), along with the deterrence outcome.

It is striking—although, on reflection, unsurprising—that the defenders in this list are predominantly major powers whereas the challengers are predominantly minor powers. For the whole data set, the short-term balance is above average for 15 of the 20 minor-major crises (75%), whereas it is above average for only 4 of the 38 remaining cases (11%).<sup>12</sup>

12. In a minor-major crisis, the challenger is a minor power and the defender a major power. I have used Singer and Small's (1972, 23) designations, which were given for 1816-1965. If these are extended to the present, the major powers since 1885 are: Austria-Hungary, 1885-1918; Britain, 1885 on; China, 1949 on; France 1885-1940, 1945 on; Germany 1885-1918, 1925-1945; Italy 1885-1943; Japan 1895-1945; Russia/Soviet Union 1885-1917, 1922 on; United States 1898 on. Were the list updated, clearly Japan and Germany should recover major power rank by some point in the 1970s or 1980s. This is immaterial for the study because neither has recently been involved in an extended immediate deterrence crisis. One could argue about Japan starting at 1895 rather than 1904; Italy being included at all; and perhaps China, Britain, and France after 1945. But changing any of these would not materially affect the results. Essentially, these designations provide a crude index of military power and reputation, which is what is needed.

In addition, note that the list in Table 1 contains a high proportion of cases in which a minor power threatens some overseas interest of a major power, about which there is room for considerable uncertainty concerning the major power's willingness to defend with force. A few examples: Iraq first overtly threatens Kuwait in 1961, shortly after Kuwaiti independence had been negotiated from Britain. Libya threatens to intervene a second time in Chad in 1983, in the wake of French military withdrawal from the country in 1979-1980 and a French decision not to resist the first Libyan advance in 1981. Turkey threatens Syria in 1957 following governmental shifts within Syria that led to increased cooperation with the Soviet Union; the extent of military support the latter would give Syria in a crisis had to be quite uncertain, given that this was a recently changed relationship. Newly elected, Kennedy's willingness to provide armed support for the Laotian government against the Pathet Lao and North Vietnam was similarly uncertain, *ex ante*.

Thinking about the cases on the list suggests that the short-term balance is separating out a set in which observably weaker challengers are making limited probes, highly uncertain about a more powerful defender's willingness to use force. If so, then the defender's immediate deterrent threat is more likely to succeed in these cases not because of *ex ante* misperception of the defender's ability to blunt a rapid offensive, but rather due to the challenger's prior expectations concerning the defender's response. The short-term balance may be proxying for initial beliefs about the defender's willingness to use force on the issue.

Examining how deterrence outcomes vary with minor-major status of challenger and defender reveals a striking pattern consistent with this argument. The set of minor-major cases—challenger is a minor power, defender is a major power—is listed along with the effect of the defender's threat in Table 2. Here we see even more clearly a set dominated by limited probes made when the challenging minor power was quite uncertain about the defender's resolve on the specific issue. For example, this list picks up two relatively opportunistic ventures by Turkey (1906 and 1922) and two by Serbia (1908 and 1912) in which the short-term balance variable is below average.<sup>13</sup>

At the same time, immediate deterrence succeeded in 18 out of the 20 minor-major conflicts (90%), as compared with 16 out of the other 34 major-major, minor-minor, and major-minor disputes (47%). This distribution would arise purely by chance in less than 1 out of 1,000 samples (*chi*

13. On the cases involving Serbia, for example, Lebow and Stein (1990a) argue that in neither did Serbian officials have any "serious intention" of going to war, but rather challenged in the hope of attracting diplomatic, and particularly Russian, support for their demands (pp. 31, 35-36).

TABLE 2  
 Extended Immediate Deterrence Crises In Which the  
 Challenger Is a Minor Power and the Defender Is a Major Power

<i>Case</i>	<i>Year</i>	<i>Challenger</i>	<i>Protégé</i>	<i>Defender</i>	<i>Outcome</i>
2	1885-1886	Bulgaria	Serbia	Austria-Hungary	Success
3	1886	Greece	Turkey	Britain	Success
5	1897	Greece	Crete	Britain/Turkey	Success
9	1903-1904	Columbia	Panama	United States	Success
12	1906	Turkey	Egypt	Britain	Success
13	1908-1909	Serbia/Russia	Austria-Hun.	Germany	Success
17	1912-1913	Serbia/Russia	Austria-Hun.	Germany	Success
18	1913	Rumania	Bulgaria	Russia	Success
20	1913	Serbia	Albania	Austria-Hungary	Success
25	1921	Panama	Costa Rica	United States	Success
26	1922	Turkey	Greece	Britain	Success
41	1957	Turkey	Syria	Soviet Union	Success
42	1961	Iraq	Kuwait	Britain	Success
43	1961	North Vietnam	Laos	United States	Success
46	1964-1965	Indonesia	Malaysia	Britain	Failure
47	1964-1965	North Vietnam	S. Vietnam	United States	Failure
52	1971	India	Kashmir (Pakistan)	China	Success
55	1975	Guatemala	Belize	Britain	Success
56	1977	Guatemala	Belize	Britain	Success
58	1983	Libya	Chad	France	Success

square is 12.4). Moreover, minor-major status is more strongly correlated with immediate deterrence success than either the long-term balance or the short-term balance and remains the only significant variable in a logistic regression with deterrence outcome as the dependent variable, including the other military balance variables as independent variables.

The empirical pattern is striking: over the past 100 years, extended immediate deterrence almost always worked when applied by a major against a minor power. To explain this pattern as the result of consistent *ex ante* misperception of the relevant military balance by minor power challengers is possible, but implausible. It requires us to believe that minor power challengers systematically fail to recognize that their major power adversaries possess superior forces and then suddenly realize this fact after the defender responds.

If this were the explanation, then we might also find a set of cases in which minor powers resist challenges by major powers. Systematic misperception should lead some minor powers to resist threats in which they subsequently

realize they will have to acquiesce. But despite the fact that major powers have much more active foreign policies than minor powers, there is only one major-minor dispute in the data set.<sup>14</sup>

Turning to the effect of nuclear weapons on immediate deterrence, we find a sharper difference between the two theories. Hypothesis 1 predicts that as an *ex ante* indicator, possession of nuclear weapons by the defender should be related to immediate deterrence success. As with the short-term balance, a nuclear-armed state will tend not to be threatened on issues known to be of major importance to it; thus we should find nuclear-armed defenders' immediate deterrent threats working more often. By contrast, Huth (1988b) originally argued that possession of nuclear weapons will not make immediate deterrence more likely to work because the nuclear threat will be incredible in most extended deterrence crises.

Huth's empirical results appeared at first glance to support his original hypothesis. When included in his "most powerful equation," a dummy variable coded 1 when the defender has nuclear weapons is not statistically significant at the .05 level used as a cut-off (Table 3).<sup>15</sup>

Note, however, that the estimated effect of having nuclear weapons is positive and respectably large. The nuclear variable was discarded essentially because the standard error of its coefficient is also somewhat large—there is a slightly less than one-in-five chance that a coefficient of this size would appear if in fact nuclear weapons had no effect on immediate deterrence.

It is not hard to see why this standard error is as large as it is: nuclear weapons existed only after 1944 and were possessed by the defender in only 15 of the 58 cases in the sample. Because the dummy variable does not vary much, logistic regression produces an unbiased but relatively uncertain estimate. In essence, in the 1988 analysis Huth discarded the nuclear weapons variable for lack of information rather than evidence that nuclear weapons do not affect immediate deterrence.

What evidence there is supports the hypothesis that there is a positive relationship.<sup>16</sup> Because nuclear weapons were simply not available before 1945, it seems reasonable to focus on the postwar cases. In this set ( $N = 24$ ),

14. And this dispute is a 1911 conflict between Italy and Turkey—one can reasonably question Italy's major power status at this time.

Minor power leaders may tend to have lower domestic political audience costs for backing out of a crisis (Fearon 1992, forthcoming-a). If so, then this may be another reason for the high rate of immediate deterrence success when applied by major against minor powers. In the model, if minor power signaling costs are low, opportunistic challenges are made worthwhile even given initial beliefs about the likelihood of resistance that would lead a state with higher audience costs to shy away.

15. The definition of several of the control variables in this equation will be discussed below.

16. As noted above, Huth reaches this same conclusion via a different route in his 1990 paper.

TABLE 3  
 Logit Equation for the Effect of the Defender's Nuclear  
 Status on Immediate Deterrence Outcome (1 = success, 0 = failure)

<i>Independent Variable</i>	<i>Coefficient</i>	<i>Standard Error</i>
Nuclear defender	1.36	1.00
Immediate balance	1.28	.63
Short-term balance	1.76	.80
Tit-for-tat	1.59	.84
Firm-but-flexible	1.85	.91
Diplomatic defeat	-2.12	1.30
Diplomatic put-down	-1.31	.88
Constant	-3.78	
<i>N</i> = 58		
Log likelihood = -22.24 (6 <i>df</i> )		

defender's nuclear status is more strongly correlated with immediate deterrence success than any other variable in Huth's preferred equation. The bivariate correlations with the dependent variable are nuclear weapons, .57; immediate balance, .08; short-term balance, .34; tit-for-tat, .07; firm-but-flexible, .44; diplomatic defeat, -.08; diplomatic put-down, -.29.<sup>17</sup>

To get a sense of what lies behind the relatively strong correlation between nuclear status and immediate deterrence, it is helpful to look more directly at the postwar cases. In these, immediate deterrence worked 80% of the time when defenders had nuclear weapons (12 of 15) and only 22% of the time when they did not (2 of 9 cases; this distribution would arise by chance in less than 1 out of 100 samples).

Table 4 lists the postwar cases and results of the defender's threat, according to the defender's nuclear status. Comparing the nuclear and non-nuclear sets reveals a fairly clear difference. When the defender had nuclear weapons, challenges tended to be probes on peripheral and often far-away interests. For example, in only 1 of 15 cases did a nuclear defender share a common border with the threatened protégé. By contrast, when the defender did not have nuclear weapons, crises more often concerned issues expected to be important to challenger and defender alike. In 5 out of the 9 cases, nonnuclear defenders were contiguous with the threatened protégé.

Diplomatic historical work on these crises would also seem to support the view that challenges to protégés of nuclear defenders were rarely made with

17. The same results can be seen in logistic regression of deterrence outcome on these variables: the dummy variable for nuclear status has the largest coefficient and the largest *t* ratio. However, because there are so few data, none of the coefficients are significant at the .10 level. (It is somewhat reassuring that the estimated effects are not very unstable.)

TABLE 4  
 Post-World War II Cases of Extended Immediate  
 Deterrence Listed by Nuclear Status of the Defender

<i>Case</i>	<i>Year</i>	<i>Challenger</i>	<i>Protégé</i>	<i>Defender</i>	<i>Outcome</i>
Defender has nuclear weapons					
35	1946	Soviet Union	Iran	United States	Success
36	1946	Soviet Union	Turkey	United States	Success
37	1948	Soviet Union	West Berlin	United States	Success
38	1950	China	Taiwan	United States	Success
40	1954-1955	China	Quemoy- Matsu	United States	Success
41	1957	Turkey	Syria	Soviet Union	Success
42	1961	Iraq	Kuwait	Britain	Success
43	1961	North Vietnam	Laos	United States	Success
46	1964-1965	Indonesia	Malaysia	Britain	Failure
47	1964-1965	North Vietnam	S. Vietnam	United States	Failure
52	1971	India	Kashmir (Pakistan)	China	Success
55	1975	Guatemala	Belize	Britain	Success
56	1977	Guatemala	Belize	Britain	Success
57	1979	China	Vietnam	Soviet Union	Failure
58	1983	Libya	Chad	France	Success
Defender does not have nuclear weapons					
39	1950	United States	North Korea	China	Failure
44	1961	India	Goa	Portugal	Failure
45	1961-1962	Indonesia	West Irian	Netherlands	Failure
48	1964-1965	United States	N. Vietnam	China	Success
49	1967	Israel	Syria	Egypt	Failure
50	1967	Turkey	Cyprus	Greece	Failure
51	1970	Syria	Jordan	Israel	Success
53	1974	Turkey	Cyprus	Greece	Failure
54	1975	Morocco	W. Sahara	Spain	Failure

much expectation of forcible resistance. For example, the first three nuclear cases are well-studied crises of the cold war—Soviet probes in the immediate postwar years (Iran 1945-1946, Turkey 1946, West Berlin 1948). Although analysts differ on whether Stalin's aims were primarily offensive or defensive, most agree that these challenges were judged relatively safe because they were on issues of uncertain and possibly low import for the (initially demobilizing) United States (for example, George and Smoke 1974 on Berlin 1948). The only nuclear cases for which evidence suggests the challenger expected military resistance to be likely are North Vietnam's threat to the

South, with the United States as defender, and possibly Guatemala's second challenge to Britain over Belize in 1977. By contrast, in at least six of the nine nonnuclear cases, historical evidence suggests the challenger began expecting resistance to be either fairly or quite likely.<sup>18</sup>

To summarize, although there are not many data, what data there are suggest rather strongly that immediate deterrent threats are more likely to work when the defender possesses nuclear weapons. In one sense this is counterintuitive. If we consider the cases in Table 4, we find at most two in which there is any possibility that a nuclear threat by the defender persuaded to the challenger to back off (Berlin and Quemoy-Matsu; Betts 1987). So how could it be that nuclear weapons affect the likelihood of immediate deterrent success? This initially intuitive argument misses the strategic consequences of the fact that challengers select themselves according to their prior beliefs about the defender's willingness to resist. Immediate deterrent threats by nuclear states are comparatively likely to work because such states are less likely to have been challenged on issues expected to be of major importance to them.

#### THE BALANCE OF INTERESTS AND THE EFFECTS OF THREATS

To gauge the effect of the balance of interests on immediate deterrence success, Huth (1988b) employs six measures. These are all *ex ante* indicators, observable by the states prior to the challenger's initial threat. Three are argued to reflect the strength of the defender's tie to the protégé state: whether defender and protégé have a military alliance; the level of arms transfers from defender to protégé; and the level of foreign trade between them. Huth hypothesizes that the presence of an alliance, more arms transfers, and more trade should each be associated with a greater chance of immediate deterrence success (pp. 44-46).

Three other measures are said to reflect the strength of the challenger's interest in the protégé: whether challenger and protégé are contiguous; the population of the protégé; and whether the protégé is a source of strategic raw materials for the challenger. Contiguity, a larger population, and source of strategic materials are all predicted to make immediate deterrence less likely to succeed because the challenger is argued in each case to have a greater interest in control of the protégé (Huth 1988b, 46-48).<sup>19</sup>

18. See citations and discussion in Fearon (1992, chap. 6).

19. At least by the arguments given in Huth (1988b), it is not clear why we should not expect that these variables would be equally valid as measures of the defender's interest in the protégé. More generally, the criteria for selecting which interest measures should apply to the defender and which to the challenger are unclear.

In contrast to the balance of capability variables, Huth's data analysis does not support prior expectations concerning the *ex ante* balance of interest variables. None of the six measures has an effect on the likelihood of immediate deterrence success that can be reliably distinguished from zero.

This contrasts with the results of Huth and Russett's (1984) first study, in which some of the relative interest variables were found to be statistically significant. Huth (1988b) conjectures that the inclusion of new indicators measuring the defender's bargaining behavior renders the *ex ante* interest variables insignificant (p. 83). The suggestion is that how the defender responds to the challenge is probably a more reliable signal of interest than an indirect measure such as trade or arms transfers. Huth's two measures of defender bargaining behavior are both strongly related to immediate deterrent success. These comprise a variable coded 1 when the defender adopted a tit-for-tat military response to the challenger, and a variable coded 1 when the defender's diplomatic posture was firm-but-flexible rather than bullying or conciliatory. Note that from the perspective developed here, these are both *ex post* measures of defender interest in the protégé (they are not known prior to the challenger's threat).

### Reassessment

The signaling theory agrees with the mainstream hypothesis concerning *ex post* measures of defender interest in the protégé: if the use of tit-for-tat military responses and/or a firm-but-flexible diplomatic posture reveals new information about the defender's level of interest, these measures should be related to immediate deterrence success, as they are in this data set (see Maoz 1983 for similar results using a different data set).

Predictions differ, however, concerning the *ex ante* measures. Whereas mainstream rationalist arguments see no reason to treat *ex ante* and *ex post* interest measures differently, the signaling theory suggests that *ex ante* measures of the defender's commitment to the protégé should be related to general deterrence success, but immediate deterrence failure (hypothesis 1). More generally, the signaling theory implies (1) any measure of the defender's interest in the protégé that predicts that the defender would actually fight should be related to immediate deterrence failure, and (2) any measure of the defender's interest that predicts the defender will not fight should be related to immediate deterrence success.

What evidence there is tends to support these two hypotheses—although, as in the case of nuclear weapons, there are not many data, and there is also one anomalous finding.



To see this, consider first that if an *ex ante* measure of the defender's level of interest in the protégé is a good one, it ought to predict whether the defender will fight on behalf of the protégé. Thus we can empirically evaluate balance of interest measures by examining whether they predict the use of force by the defender in the set of cases where immediate deterrence failed. For this set, Table 5 reports bivariate correlations between the defender's decision to fight (1 = fight, 0 = back down) and Huth's three measures of the defender's level of interest in the protégé (presence of an alliance, level of trade, level of arms transfers). As in Huth and Russett (1988), I have also included a fourth measure, called contiguity, which is coded 1 when defender and protégé are contiguous states or are separated by a short stretch of water and is coded 0 otherwise.

With the notable exception of arms transfers, the measures are positively correlated with the defender's decision to fight, as one would expect and hope. The signaling theory predicts that these signs should all reverse when the same variables are correlated with the immediate deterrence outcome (the challenger's decision to back off or press ahead with its threat). An *ex ante* measure that predicts the defender's decision to fight (not fight) should be related to immediate deterrence failure (success).

Table 6 reports these correlations between the four measures of defender interest in the protégé and immediate deterrence outcome (1 = success, 0 = failure). The predicted reversals occur for three of the variables (alliance, arms, and contiguity), whereas the sign moves toward zero for the fourth (trade). Excepting the trade variable, these signs remain the same in the fully specified logit model with immediate deterrence outcome as the dependent variable (Table 7).

To summarize these results verbally, the predicted sign reversals occur for three of the four measures. Two of these three interest measures are positively related to the defender's decision to fight but negatively affect immediate deterrence. Immediate deterrence is significantly less likely to work when protégé and defender have an alliance, and the same appears true if defender and protégé are geographically close, although the level of statistical confidence is lower.

The third variable, arms transfers to the protégé, is related to immediate deterrence success at the same time as it predicts that the defender will not fight. This is as anticipated by the theory: if defenders are less likely to be willing to resist with force in the cases where arms transfers are relatively high, then we should observe more opportunistic challenges in these cases (i.e., challengers with lower values for  $w_c$  on average), and thus more successful immediate deterrent threats. The puzzle is why higher arms transfers are related to decisions not to fight for the protégé. The small number

TABLE 5  
Correlation of Interest Measures with  
Defender's Decision to Fight for the Protégé ( $N = 24$ )

<i>Measure of Defender Interest in Protégé</i>	<i>R (with decision to fight)</i>
Alliance (1 = alliance, 0 = none)	.17
Contiguity (1 = close, 0 = not)	.44
Trade (percentage of all trade for defender)	.45
Arms (percentage of all delivered to protégé)	-.20

TABLE 6  
Correlation of Interest Measures  
with Immediate Deterrence Outcomes ( $N = 58$ )

<i>Defender Interest in Protégé</i>	<i>R (with immediate deterrence success)</i>
Alliance (1 = alliance, 0 = none)	-.12
Contiguity (1 = close, 0 = not)	-.17
Trade (percentage of all trade for defender)	.02
Arms (percentage of all delivered to protégé)	.14

of cases makes speculation hazardous, but it could be that arms transfers sometimes act as a low-cost substitute for a more serious commitment in cases where, for other reasons, the defender would be reluctant actually to fight. For example, the Portuguese decision not to fight India for Goa in 1961, the Spanish decision not to support the break-away Saharan republic against Morocco in 1975, and the Soviet decision not to fight with Vietnam against China in 1979 may fit this pattern (in each instance prior arms transfer levels were quite high).

Finally, the fact that the level of trade between defender and protégé predicts both the decision to fight and immediate deterrence success is anomalous from the perspective of the signaling model. If not a statistical accident, this finding appears as evidence of systematic misperception by challenging states: they apparently are not making use of the fact that trade ties between defender and protégé predict that the defender will in fact support the protégé militarily.<sup>20</sup>

20. My only reservation about this interpretation is that all of these measures (alliance, trade, arms transfers, contiguity) are incredibly crude and noisy compared to the specific *ex ante* information available to state leaders in each particular case. I would expect that such specific information would in every case allow a sharper estimate of the defender's response than would a regression formula using these crude measures. Thus, even though I would not be at all surprised if systematic misperception along these lines did occur, I would be surprised if it were picked up by such crude measures.

TABLE 7  
 Logit Equation for the Effect of Balance of Interest Variables  
 on Immediate Deterrence Outcome (1 = success, 0 = failure) ( $N = 58$ )

<i>Independent Variable</i>	<i>Coefficient</i>	<i>Standard Error</i>
Variables measuring defender interest in protégé		
Alliance	-7.27	3.33
Contiguity	-2.11	1.75
Trade	.98	.45
Arms	.41	.19
Control variables		
Immediate balance	.61	.76
Minor-major	4.32	1.84
Tit-for-tat	4.41	2.17
Firm-but-flexible	3.09	1.43
Past crisis	-2.72	1.34
Constant	-2.80	
Log likelihood = -17.07 (9 <i>df</i> )		

#### BARGAINING BEHAVIOR OF THE DEFENDER IN PRIOR CRISES

Huth's (1988b) final set of independent variables describes the defender's bargaining behavior in past confrontations, which are used to address an interesting and much debated question about the role of a state's bargaining reputation in crises: do challengers draw any significant inferences about the defender's resolve from how the defender acted in previous confrontations with the challenger or with other states?

Huth (1988b) uses three dummy variables to measure the defender's behavior in past crises. One is coded 1 if the defender bullied its earlier opponent into backing down (diplomatic put-down). A second is coded 1 if the defender "retreat[ed] under pressure and conced[ed] on the critical issues in order to avoid a direct military confrontation" (diplomatic defeat). A third is coded 1 if the defender and challenger reached a stalemate, in which "firm-but-flexible bargaining by the defender [failed] to reach an agreement which resolves the underlying issues" (pp. 69-70). Each of these is coded for both the defender's last confrontation with any challenger and for its last confrontation with the current challenger, if one occurred.

Huth offers two hypotheses concerning these variables. First, following Jervis's (1976, 239) argument that state leaders are more influenced by events that they experience firsthand, he proposes that "The past behavior of the defender in confrontations in which the current attacker was directly involved

will have a greater impact on deterrence outcomes than in cases in which the current attacker was not directly involved" (1988b, 55).

Second, concerning past encounters with the current challenger, Huth argues that diplomatic put-downs and diplomatic defeats (defender bullied or caved in) should be related to immediate deterrence failure, whereas stalemates produced by firm-but-flexible bargaining should be related to successful immediate deterrence in the later crisis. For put-downs, the idea is that the challenger will be left unhappy with the status quo, and determined to protect its bargaining reputation and avoid another retreat before the defender. For diplomatic defeats, Huth (1988b) argues that, as apparently happened for Britain and France after Munich, backing down "weakens the defender's future credibility," making the challenger more likely to disregard future threats. For stalemates, Huth simply conjectures that a record of firm-but-flexible bargaining will stand the defender in good stead in a subsequent crisis (p. 55).

In sum, Huth argues that variables measuring the defender's past bargaining behavior will allow at least a partial empirical test of a contentious issue in the field of security studies—how important is reputation? He hypothesizes that reputational considerations will matter less than Schelling (1966) and others have suggested, but will nonetheless play a role in determining both the motivation of challenger and the inferences it draws from a defender's threats.

On balance, the empirical results appear to support the hypotheses. Past behavior matters only when the earlier conflict involved the current challenger. And when the defender had either bullied the challenger or caved in to its demands, immediate deterrence was significantly less likely to work. However, the data provide no support for the hypotheses that stalemates and a record of firm-but-flexible bargaining will improve one's chances for dissuading a challenger in subsequent encounters.

## Reassessment

A closer look at how these results emerge calls this interpretation into doubt. Consider the proposition that how the defender bargained in its most recent conflict with the current challenger matters. Huth assesses this by looking at the relationship between immediate deterrence outcomes and the dummy variable for each of the three broad styles of bargaining and crisis outcomes he identifies: diplomatic put-down (bullying); diplomatic defeat (highly conciliatory); and stalemate (firm-but-flexible bargaining by the defender). Each variable is coded 1 if the defender used this style in a previous crisis with the same challenger. It is coded 0 if either (1) there was a past confrontation but the defender used one of the other styles, or (2) there was

no past confrontation between the challenger and defender in question. Thus there is a set of cases in the data set that are coded 0 for all three dummy variables—these are the cases for which there was no previous confrontation between defender and challenger.

When the three variables are included with Huth's other preferred independent variables, the logit results reported in Table 8 appear. Note that all three bargaining styles/past crisis outcomes have a negative impact on immediate deterrence (although the estimated effect of stalemate is only a bit more than one standard deviation from zero). The reason becomes clear with a little thought. The effect of these variables is negative relative to the cases in the excluded category—that is, relative to the cases in which there was no previous crisis between defender and challenger. It follows that the correct interpretation of the significant negative coefficients for diplomatic put-down and diplomatic defeat is not that how the defender bargained or the outcome of the last crisis matters. Rather, these coefficients indicate simply that, compared to cases with no prior conflict, immediate deterrence was less likely to work when there was a prior crisis and one of these bargaining styles/outcomes obtained.

The appropriate way to test if the defender's bargaining style in the last crisis matters is to ask whether there are significant differences between the coefficients for the three dummy variables. Do the data allow us to reject the hypothesis that bullying, conciliatory, or firm-but-flexible strategies all have the same effect on the chance that the defender's threat will work in the current crisis? The appropriate statistical test is an *F* test of the null hypothesis that the three coefficients are the same. When applied, this indicates that there is about a one-in-three chance that coefficients this different would appear if the null hypothesis were correct. So the data really do not allow us to reject the possibility that it does not matter how the defender bargained or what outcome resulted in its last confrontation. It may matter, but the data do not provide firm evidence.

What the data do suggest is that it matters whether the defender and challenger had a previous crisis encounter, consistent with hypothesis 3 above. Immediate deterrence failed in only 2 of the 18 cases where there was no prior crisis between defender and challenger (11%), as opposed to 18 of the 40 cases with a history (45%). If we construct a new independent variable—past crisis, coded 1 when there was a prior crisis between defender and challenger, and 0 otherwise—we find that it has a significant independent effect on immediate deterrence failure, as seen in the logistic regression reported in Table 9.<sup>21</sup>

21. The results do not change if we include all of the independent variables for which this analysis has found a theoretical and/or empirical rationale for inclusion (see Fearon 1992, chap. 6).

TABLE 8  
 Logit Equation for the Effect of Defender's Past Behavior  
 on Immediate Deterrence Outcome (1 = success, 0 = failure) ( $N = 58$ )

<i>Independent Variable</i>	<i>Coefficient</i>	<i>Standard Error</i>
Variables measuring past bargaining behavior		
Diplomatic put-down	-2.32	1.12
Stalemate	-1.21	1.09
Diplomatic defeat	-3.01	1.44
Control variables		
Immediate balance	1.08	.54
Short-term balance	1.59	.77
Tit-for-tat	1.56	.83
Firm-but-flexible	1.97	.91
Constant	-2.34	
Log likelihood = -22.57 (7 df)		

This empirical pattern is more plausibly explained by selection effects produced by strategic behavior than by considerations of bargaining reputation. The fact that the defender was willing to resist or challenge on the same (or a related) issue in the past is an *ex ante* indicator of the defender's willingness to use force (relatively high  $w_D$  in the model). As such, a previous crisis implies, for a subsequent crisis, a greater initial belief that the defender might resist, and hence a lower chance immediate deterrence will succeed.

What is slightly counterintuitive here is that regardless of how the defender acted or the outcome of the previous crisis, challengers' prior beliefs should on average be less optimistic in the set of cases with previous conflicts than in the set without. Thus, even when the defender "caved in" in the previous crisis, any subsequent challenge will be made with an initial belief higher (on average) than for cases with no prior crisis. This is a consequence of costly signaling and the properties of equilibrium in the model. The fact that a state was willing to try resistance, even if it ultimately made concessions, reveals a higher willingness to use force on the issue than if the state had not resisted at all (and no crisis had occurred). In other words, regardless of the outcome, the fact that a past crisis occurred suggests that there is at least one issue on which both states have higher than normal values for conflict. In turn, this implies that a subsequent challenge will be made with a higher initial belief about the defender's willingness to resist, which implies that immediate deterrence will be less likely to work.<sup>22</sup>

22. Again, this argument should be checked with a repeated version of the crisis game.

TABLE 9  
 Logit Equation for Effect of Defender's Past Behavior on Immediate Deterrence (1 = success, 0 = failure): An Alternative Interpretation ( $N = 58$ )

<i>Independent Variable</i>	<i>Coefficient</i>	<i>Standard Error</i>
Past crisis	-1.91	.97
Immediate balance	.86	.48
Minor-major	2.29	1.00
Tit-for-tat	1.56	.77
Firm-but-flexible	1.38	.77
Constant	-1.04	
Log likelihood = -23.47 (5 <i>df</i> )		

In sum, the empirical relationship observed in the cases is not between immediate deterrence and the defender's bargaining behavior in its last crisis with the current challenger. Although it may be true, the data do not provide significant support for the hypothesis that bullying the challenger makes the state more eager to defend its "bargaining reputation" the next time around, or that being highly conciliatory leads the challenger to discredit the defender's threats in subsequent crises. Rather, the empirical relationship measured by the past behavior variables is between the fact of there having been a previous crisis and subsequent immediate deterrence failure. Strategic dynamics arising from costly signaling may explain this relationship.

This argument casts a somewhat different light on the remaining finding—that immediate deterrence is unaffected by how the defender bargained in past crises with states other than the current challenger. This result does indeed undercut a strong claim about challengers' perceptions of the interdependence of commitments. If potential challengers drew strong inferences about an adversary's resolve from its behavior in unrelated conflicts, then we might well expect an effect of such behavior on immediate deterrence.

However, if there is no clear link between immediate deterrence and how the defender bargained in a past crisis with the same challenger, then it is not at all surprising that crises with other states have no apparent influence. Based on the preceding data analysis, I would argue that the key to understanding variation across cases in immediate deterrence outcomes is each side's prior expectations about the importance to the other side of the issues at stake. Relative power variables appear to matter, for example, but their effects are mediated by strategic choices concerning the issues in dispute. If most of the action concerns the nature of the issues between challenger and defender, then even if states do sometimes draw inferences about an adversary by its behavior in other confrontations, such behavior may have little value for

explaining variation in immediate deterrence outcomes in an aggregate analysis.

## CONCLUSION

This article has examined the question of how relative military capabilities and relative political interests influence the efficacy of threats made during international disputes. I have contrasted some widely accepted rationalist claims—namely, threats are more likely to work when the threatener is favored by the balance of capabilities or interests—with the results of a strategically richer game-theoretic analysis. I hope to have shown how this second approach provides more compelling explanations for observed empirical regularities and also enables the discovery and explanation of new patterns in the data.

To summarize, Huth's (1988b) interpretation of evidence, guided for the most part by mainstream balance of capabilities and balance of interests arguments, yielded the following conclusions:

1. Immediate deterrent threats are more likely to work when the defender has the ability to blunt a rapid offensive attack.
2. Possession of nuclear weapons has no influence on immediate deterrence.
3. Ex ante indicators of the defender's level of interest in a threatened protégé state do not appear to influence positively the prospects for immediate deterrence, contrary to expectations.
4. Challenger and defender bargaining reputations, established in their most recent past confrontation, influence the prospects for immediate deterrence success.

When the data are analyzed in light of the equilibrium results sketched in the second section, a different set of conclusions emerges:

1. Although the capacity to blunt a rapid offensive may make general deterrence more likely to succeed, its apparent effect on immediate deterrence is not due to the specific military attributes of the balance. Rather, when the defender is relatively strong, challengers tend to threaten on issues on which a concerted response by the defender is quite uncertain; hence a costly signal in response is likely to work.
2. Possession of nuclear weapons does appear to be related to immediate deterrence success, not because challengers have feared nuclear escalation but again because of selection effects. States are unlikely to challenge nuclear powers on issues perceived to be important to them. Instead, challengers will tend to threaten nuclear powers over issues on which a concerted response by the defender is initially quite uncertain, and hence a costly signal in reply is more likely to work.



3. Ex ante measures of the balance of interests such as alliance ties or geographical contiguity between defender and protégé appear to be related to the failure of immediate deterrent threats. This is contrary to the predictions of standard balance of interests accounts, but as predicted by the costly signaling model.
4. Although bargaining reputations acquired in past confrontations may influence the likelihood that a state's immediate deterrent threats will work, the data do not allow rejection of the contrary proposition that they do not matter. Instead, the evidence indicates that immediate deterrent threats are less likely to work when there was some previous crisis between challenger and defender. This finding is both predicted and explained by the equilibrium results of the crisis game discussed above.

I conclude with two broader comments. First, the problem with the mainstream rationalist arguments about the impact of relative capabilities and interests is not that the arguments are flat-out wrong. There are situations in which an advantage in military capabilities of "intrinsic interests" should be correlated with the success of a state's threats or warnings. For example, in the model described above, this is exactly what we would expect for general deterrent threats. Problems arise when we attempt to apply the mainstream hypothesis to specific strategic contexts—such as threats made during an international dispute—without thinking about their potentially distinctive features.

Second, the analysis suggests that empirical studies of political matters that involve strategic behavior should be guided, or at least informed, by developed theoretical analyses of the strategic problems faced by the agents. Both the construction of data sets and the interpretation of empirical findings tend to be strongly shaped by the implicit or explicit theoretical apparatus employed by the analyst. Making the theoretical analysis explicit, particularly in game-theoretic form, may suggest problems with conventional informal arguments and provide a new understanding of what to look for and how to interpret empirical results.

## REFERENCES

- Betts, R. 1987. *Nuclear blackmail and nuclear balance*. Washington, DC: Brookings Institution.
- Brecher, M., J. Wilkenfeld, and S. Moser. 1988. *Handbook of international crises*. New York: Pergamon.
- Bueno de Mesquita, B., and D. Lalman. 1992. *War and reason*. New Haven, CT: Yale University Press.
- Fearon, J. 1990. Deterrence and the spiral model: The role of costly signals in crisis bargaining. Paper presented at the annual meetings of the American Political Science Association, San Francisco.

- . 1992. Threats to use force: The role of costly signals in international crises. Ph.D. diss., University of California, Berkeley.
- . Forthcoming-a. Domestic political audiences and the escalation of international disputes. *American Political Science Review*.
- . Forthcoming-b. Selection effects and deterrence. In *Deterrence debates*, edited by K. Oye. Ann Arbor: University of Michigan Press.
- George, A. L., and R. Smoke. 1974. *Deterrence in American foreign policy*. New York: Columbia University Press.
- Gochman, C., and Z. Maoz. 1984. Militarized interstate disputes, 1816-1976: Procedures, patterns, insights. *Journal of Conflict Resolution* 28:585-615.
- Huth, P. 1988a. Extended deterrence and the outbreak of war. *American Political Science Review* 82:423-44.
- . 1988b. *Extended deterrence and the prevention of war*. New Haven, CT: Yale University Press.
- . 1990. The extended deterrent value of nuclear weapons. *Journal of Conflict Resolution* 34:270-90.
- Huth, P., and B. Russett. 1984. What makes deterrence work? Cases from 1900 to 1980. *World Politics* 36:496-526.
- . 1988. Deterrence failure and crisis escalation. *International Studies Quarterly* 32:29-46.
- . 1993. General deterrence between enduring rivals: Testing three competing models. *American Political Science Review* 87:61-73.
- Jervis, R. 1971. Bargaining and bargaining tactics. In *Nomos: Coercion*, edited by J. Roland Pennock and J. Chapman, 272-88. Chicago: Aldine.
- . 1976. *Perception and misperception in international politics*. Princeton, NJ: Princeton University Press.
- . 1989. Rational deterrence: Theory and evidence. *World Politics* 42:183-207.
- Kilgour, D. M. 1991. Domestic political structure and war behavior: A game-theoretic approach. *Journal of Conflict Resolution* 35:266-84.
- Lebow, R. N., and J. G. Stein. 1990a. The 1984 Huth and Russett data set. Typescript, Cornell University.
- . 1990b. *When does deterrence succeed and how do we know?* Ottawa: Canadian Institute for International Peace and Security.
- Leng, R. 1993. *Interstate crisis behavior, 1816-1980: Realism versus reciprocity*. Cambridge: Cambridge University Press.
- Maoz, Z. 1983. Resolve, capabilities, and the outcomes of interstate disputes, 1816-1976. *Journal of Conflict Resolution* 27:195-229.
- Maxwell, S. 1968. *Rationality in deterrence*. London: International Institute for Strategic Studies.
- Mearsheimer, J. 1983. *Conventional deterrence*. Ithaca, NY: Cornell University Press.
- Morgan, P. 1977. *Deterrence: A conceptual analysis*. Beverly Hills, CA: Sage.
- Morrow, J. D. 1989a. Bargaining in repeated crises: A limited information model. In *Models of strategic choice in politics*, edited by P. C. Ordeshook, 207-28. Ann Arbor: University of Michigan Press.
- . 1989b. Capabilities, uncertainty, and resolve: A limited information model of crisis bargaining. *American Journal of Political Science* 33:941-72.
- Nalebuff, B. 1986. Brinkmanship and nuclear deterrence: The neutrality of escalation. *Conflict Management and Peace Science* 9:19-30.

- Osgood, R., and R. Tucker. 1967. *Force, order, and justice*. Baltimore, MD: Johns Hopkins University Press.
- Pape, R. 1990. Coercive air power in the Vietnam War. *International Security* 15:103-46.
- Posen, B. 1984. *The sources of military doctrine*. Ithaca, NY: Cornell University Press.
- Powell, R. 1990. *Nuclear deterrence theory: The problem of credibility*. Cambridge: Cambridge University Press.
- Schelling, T. 1966. *Arms and influence*. New Haven, CT: Yale University Press.
- Shimshoni, J. 1988. *Israel and conventional deterrence: Border warfare from 1953 to 1970*. Ithaca, NY: Cornell University Press.
- Singer, J. D., and M. Small. 1972. *The wages of war, 1816-1965: A statistical handbook*. New York: Wiley.
- Snyder, G. H., and P. Diesing. 1977. *Conflict among nations: Bargaining, decision making, and system structure in international crises*. Princeton, NJ: Princeton University Press.
- Snyder, J. 1984. *The ideology of the offensive: Military decision making and the disaster of 1914*. Ithaca, NY: Cornell University Press.
- Spence, A. M. 1973. Job market signalling. *Quarterly Journal of Economics* 87:355-74.
- Van Evera, S. 1984. Causes of war. Ph.D. diss., University of California, Berkeley.
- Wagner, R. H. 1991. Nuclear deterrence, counterforce strategies, and the incentive to strike first. *American Political Science Review* 85:727-49.