International Studies Quarterly (1993) **37**, 1–27

Asymmetric Deterrence

FRANK C. ZAGARE

State University of New York at Buffalo

AND

D. MARC KILGOUR

Wilfrid Laurier University

Deterrence of a challenger by a defender is modeled by explicitly relating uncertainty and the credibility of retaliatory threats to the stability of an asymmetric deterrence relationship. In the two-person game model, each player either prefers to fight rather than back down or prefers the reverse. A player knows its own preference, but is uncertain of its adversary's. The challenger may choose to accept the status quo or initiate a crisis; in the latter case, the defender may capitulate or defend; if it defends, the challenger must either back down or face a situation of open conflict. The perfect Bayesian equilibria of the game are determined, interpreted, and illustrated with historical examples of the success or failure of deterrence.

In this essay we explore a simple game-theoretic model of asymmetric deterrence. Specifically, we posit a game in which one decision-maker, "Challenger," decides whether to initiate a crisis involving a second decision-maker, "Defender." We assume this decision is made without complete information about the relative preferences of Defender for holding firm (resisting) or capitulating (backing down). In turn, Defender must formulate its response without full information about Challenger's preferences. Thus, the conclusions we draw from the analysis of this game augment our understanding of the connection between certain kinds of beliefs and deterrence stability. They also shed light on the role played by credible threats in averting acute international crises.

Asymmetry and Deterrence

The foundations of modern deterrence theory were laid against the backdrop of the Cold War. Probably because most theorists of that era were preoccupied with the question of how the Soviet Union might be deterred from attacking Western

Authors' note: This material is based upon work supported by the National Science Foundation under Grant No. SES-9123219 to Frank C. Zagare. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. D. Marc Kilgour gratefully acknowledges the support of the Laurier Centre for Military, Strategic, and Disarmament Studies, the Laurier Research Professorship, and the Natural Sciences and Engineering Research Council of Canada.

Published by Blackwell Publishers, 238 Main Street, Cambridge, MA 02142, USA, and 108 Cowley Road, Oxford OX4 1JF, UK.

interests (and not vice versa), the—usually unstated—assumption of asymmetry in offensive motivation figures prominently in the strategic literature. Daniel Ellsberg's (1959:358–359) seminal discussion of the calculus of deterrence, for example, poses deterrence as a fundamentally one-sided problem: how to deter a blackmailer via threats when the cost of executing threats is prohibitive.

The theoretical distinction between asymmetric and symmetric deterrence is also implicit in Morgan's (1977:28) well-known definition of "immediate" deterrence as a situation in which "one side is seriously considering an attack while the other is mounting a threat to prevent it." Such deterrence situations are almost always asymmetric. Rarely do two states simultaneously plan to attack one another, save for those circumstances in which one state expects the other to attack first and so launches a preemptive war (Levy, 1987).

Likewise, asymmetric deterrence is presupposed in the literature of acute international crises. The standard definition depicts a crisis as a situation characterized, *inter alia*, by shortness of decision time and strategic surprise (Hermann, 1969). Such conditions are likely to be satisfied only when one state has already directly challenged the interests of another, suggesting once again differences in motivation and circumstances between the states involved in the crisis. Snyder and Diesing's (1977) now classic description of the anatomy of a crisis is unambiguous: crises are the direct result of a decision by one state to challenge the security interests of another.

As well, the literature of interstate war often posits an asymmetric, hierarchically structured international system. Whereas most realists, and almost all balance-of-power theorists, view states as undifferentiated power-maximizers (Waltz, 1979), system-level theories—such as the theory of the long cycle (Modelski, 1983; Modelski and Thompson, 1989), the power cycle theory (Doran, 1989a, 1989b), the hierarchical equilibrium model (Midlarsky, 1988), and the theory of hegemonic stability (Kindleberger, 1974, 1976; Gilpin, 1975, 1981; and Krasner, 1976)—distinguish among the positions occupied by, and the situations faced by, states in the system. Organski and Kugler's (1980) power transition theory explicitly envisions an environment that may disintegrate as a consequence of a challenge made by a dissatisfied state against a status quo power defending an institutionalized order.[1]

Given the tenor of the times in which they were developed, it is hardly surprising that many influential (Western) theories of interstate conflict rest, either implicitly or explicitly, on the premise of asymmetry. Still, it would be difficult to maintain that the recent dramatic world events have rendered these theories, or the problems they address, obsolete. The end of the bipolar world and the era of superpower parity does not necessarily portend an era of peace and international stability (Mearsheimer, 1990). As Kissinger (1992) suggests, "a challenge [to the existing order] could evolve from chaos on the territory of the former Soviet Union, from ethnic conflicts and political instability in Eastern Europe, and from the re-definition of Germany's role."

As old rivalries are renewed, attempts will undoubtedly be made to settle old scores; and as new conflicts develop, the status quo will assuredly be challenged. Some of these confrontations may be multisided challenges, involving several unsatisfied parties. But more common, almost surely, will be asymmetrical conflicts—as those that have try to protect what they have, and those that have not try to procure it.

[1]For the connection between power transition theory and the game examined below, see Kugler and Zagare (1990) and Zagare (1992). Also see Kim and Morrow (1990).

Asymmetric Deterrence

Because the success or failure of asymmetric deterrence relationships is likely to remain important as the international system evolves, we propose to explore their underlying properties game-theoretically. In our asymmetric deterrence game (see Figure 1), the first player, Challenger, must choose between initiating or not initiating a challenge. If Challenger does not initiate, the game ends at the status quo—outcome CC. If a challenge is offered, and a crisis precipitated, the second player, Defender, must decide whether or not to defend its interests. If Defender gives in, the game ends at outcome DC and Challenger gains an advantage. But if Defender resists, Challenger must then decide whether to back down—in which case Defender gains an advantage[2] (outcome CD)—or face a conflict—in which case there is confrontation and possible war (outcome DD).

For convenience, the four possible outcomes of the asymmetric deterrence model, as well as the notation for the (von Neumann-Morgenstern utility) payoffs associated with them, are summarized in Figure 2. Payoffs are represented as an ordered pair in each cell of the matrix, the first entry being Challenger's utility, and the second, Defender's. We assume, with no loss of generality, that Challenger receives 1 at DC, c_3 at CC, 0 at CD, and C_2 at DD. Similarly, Defender receives 1 at CD, d_3 at CC, 0 at DC, and D_2 at DD.

To add some real-world structure to the asymmetric deterrence game, we place some additional restrictions on the players' utilities. First, we assume Challenger strictly prefers an advantage to the status quo. This axiom is necessary since, without it, deterrence is spurious. On the other hand, we place no such restriction on the preferences of Defender, who could prefer either maintaining the status quo or facing down a challenge. (It turns out that Defender's relative evaluation of the status quo and an uncontested victory is without strategic significance.) Thus we stipulate that $0 < c_3 < 1$ and $d_3 > 0$.

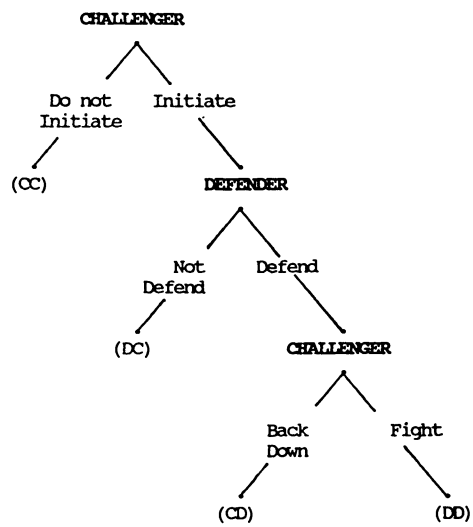


FIG. 1. Asymmetric deterrence.

[2]We make this assumption lest Challenger be faced with a totally costless decision. This is not to say that the costs are necessarily large.

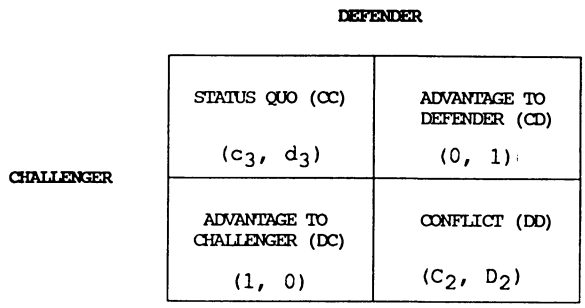


FIG. 2. Outcomes and preference notation of asymmetric deterrence games.

Second, we assume that each player prefers either winning without a confrontation, CD or DC, or the status quo, CC, to conflict, DD. [Formally, $C_2 < c_3$, and $D_2 < \min\{d_3, 1\}$.] This constraint means that a confrontation is certain to be damaging to both players. In other words, each player's threat against the other is *capable*. We make this assumption simply because deterrence is not possible without it (Zagare, 1987:ch. 4).

We also assume each player knows the other's preferences satisfy these restrictions but that each player is uncertain about whether the opponent prefers confrontation, DD, to capitulation, CD or DC. To model this uncertainty, we treat the players' payoffs at DD, C_2 (Challenger), and D_2 (Defender) as independent, binary random variables (which we denote by uppercase letters) with known distributions. More specifically, it is common knowledge that

$$C_2 = \begin{cases} c_2^+ \text{ with probability } p_{\text{Ch}} \\ c_2^- \text{ with probability } 1 - p_{\text{Ch}} \end{cases}$$

$$D_2 = \begin{cases} d_2^+ \text{ with probability } p_{\text{Def}} \\ d_2^- \text{ with probability } 1 - p_{\text{Def}} \end{cases}$$

so that both players know these distributions, both know they know, and so on. However, only Challenger knows the actual value of C_2, and only Defender knows the actual value of D_2. (The interpretations of c_2^+ and c_2^-, and d_2^+ and d_2^-, will be discussed below.) The complete set of restrictions satisfied by the eight fixed parameters can now be summarized.

$$\begin{aligned} & 0 \le p_{\text{Ch}} \le 1, \\ & 0 \le p_{\text{Def}} \le 1, \\ c_2^- < \; & 0 < c_2^+ < c_3 < 1, \\ d_2^- < \; & 0 < d_2^+ < \min\{d_3, 1\}. \end{aligned}$$

One way to think about the utility values associated with the conflict outcome is in terms of the *credibility* of each player's deterrent threat. The conflict outcome is the threat upon which a deterrence relationship rests. To deter an opponent, a player threatens to enforce mutual punishment, DD, rather than accept the outcome associated with the other's advantage, CD or DC.

This threat can be either credible or incredible. In the literature of deterrence, credible threats are generally taken to be synonymous with believable threats (Smoke, 1987). Believability, in turn, is normally equated with rationality (Lebow,

1981:15; Betts, 1987:12). Thus, only threats that would be rational to carry out are seen as credible (Kilgour and Zagare, 1991).[3]

In decision theory, rational choice is *minimally* defined as choice consistent with a player's preferences (Luce and Raiffa, 1957:50). Thus, rational threats are threats of bringing about an outcome more preferred by the threatener. In our model, if the payoff to Defender from confrontation is $D_2 = d_2^+$, then Defender's threat is rational, and hence credible—because d_2^+, Defender's value for conflict, exceeds 0, Defender's value for Challenger's advantage (i.e., DC). Players with rational threats who can resolutely enforce mutual punishment are called **Hard**. By contrast, if $D_2 = d_2^-$, Defender's threat lacks credibility because $d_2^- < 0$. This type of player, who would rather capitulate than endure mutual punishment, is called **Soft**.

In this model, therefore, the credibility of each player's threat to retaliate is measured by p_{Ch} and p_{Def}, the *a priori* probabilities that the players' threats are rational. Overall, Challenger's [respectively, Defender's] threat will be credible to the degree p_{Ch} [p_{Def}], and incredible to the degree $1 - p_{Ch}$ [$1 - p_{Def}$].

Parenthetically, it is important to point out that by defining credibility in terms of beliefs, we are able to explore the entire range of asymmetric relationships over which immediate deterrence may succeed or fail. Recent attempts to apply the methodology of games of incomplete information to the deterrence problem have done so within the confines of pre-specified preference assumptions.[4] Powell's (1990) recent models are a good example. These models, which rest upon a fixed behavioral assumption about each player's choice at the last stage of the game, postulate a constant preference for capitulation over conflict. Whereas this assumption is plausible, especially in the Nuclear Age, it is a special case of the more general relationships which we explore here.

Whereas the work of Powell explicitly specifies nations' choices in the playing out of a crisis, the models of Fearon (1990), Morrow (1989a, 1989b), and Banks (1990) focus on bargaining which may take place after the crisis has been initiated but before hostilities decide it.[5] In contrast, our approach concerns only crisis initiation, and the onset of confrontation. As well, our assumptions on preferences do not restrict generality in any way. That way functional relations among the players' utilities play no role in our conclusions.

Complete Information and Deterrence Stability

Under what conditions will Challenger initiate a crisis? When will Defender resist? When will a conflict result? What is the precise connection between deterrence and

[3]Our use of the term *rationality* is technical. Specifically, we assume "instrumental" rather than "procedural" rationality, that is, we postulate an (instrumental) link between ends and means and do not speak to the issue of what constitutes a proper (or procedurally rational) end. For an extended discussion of this point, see Zagare (1990).

[4]A good collection of articles of this genre can be found in Ordeshook (1989). For other incomplete information models of deterrence, see Banks (1990); Bueno de Mesquita and Lalman (1992); Fearon (1990); Morrow (1989a, 1989b); Nalebuff (1986); Kilgour and Zagare (1991, 1992a, 1992b); Powell (1990); and Wagner (1991). Also see Bueno de Mesquita and Morrow (1991).

[5]In Fearon's innovative study, Challenger and Defender, before deciding whether to fight, are afforded an opportunity to make a choice, which raises each player's cost of backing down and, possibly, its "credibility." Fearon's results can be seen as complementing our analysis. For a more detailed discussion of the differences between Fearon's, Morrow's, and Powell's models and the model developed here, see Kilgour and Zagare (1992a).

threat credibility? We next address these and related questions,[6] first in the simple case of complete information about the preferences of the two players, and then in the more complex setting of incomplete information.[7] Significantly, the latter exercise will enable us to explore the connection between certain kinds of belief systems and the robustness of asymmetric deterrence relationships.

When information is complete, players have accurate and full information about each other's preferences. Because we define credible (*i. e.*, rational) threats in terms of a player's preference for executing the threat, then, with complete information, Challenger's threat is either perfectly credible (when $p_{Ch} = 1$) or perfectly incredible (when $p_{Ch} = 0$), and similarly for Defender's threat. Both players know for sure whether the other is Hard or Soft, that is, whether C_2 equals c_2^+ or c_2^- and whether D_2 equals d_2^+ or d_2^-, so that credibility is common knowledge.

Given the restrictions above on the preferences of the players, there are exactly four strategically distinct asymmetric deterrence games of complete information: (1) Both players have a credible threat (*i.e.*, $p_{Ch} = 1$ and $p_{Def} = 1$); (2) only Challenger's threat is credible ($p_{Ch} = 1$ and $p_{Def} = 0$); (3) only Defender's threat is credible ($p_{Ch} = 0$ and $p_{Def} = 1$); or (4) neither player has a credible threat ($p_{Ch} = 0$ and $p_{Def} = 0$).

Table 1 summarizes the distinguishing characteristics of these four games. To display conveniently each player's preferences, the outcomes of each game are rank-ordered from best, "4", to worst, "1". Note that a player's threat is credible when, and only when, that player strictly prefers conflict, DD, to the opponent's advantage, DC or CD. This occurs only when the player gives rank "2" to DD.

Using backward induction, it is easy to determine the game-theoretic outcomes of these four games. To illustrate, the first game of Table 1 is represented in extensive form in Figure 3.[8] In this game each player's threat is credible. As the arrows in Figure 3 (representing rational choice) reveal, and as Table 1 reports, the status quo CC is the equilibrium outcome of the game. Thus, as one would expect, when each player has a credible threat, deterrence succeeds. Interestingly, as Table 1 indicates, deterrence also succeeds when only Defender's threat is credible, and when neither player has a credible threat. Defender's credibility is, therefore, not a necessary condition for deterrence, but it is sufficient.

By contrast, deterrence fails when only Challenger's threat is credible. Consequently, our model indicates that a defender without a credible threat is a necessary, though not sufficient, condition for a deterrence failure under conditions of complete information.[9] Parenthetically, we note that, according to our model, nothing like an immediate deterrence situation (Morgan, 1977:ch. 1) can emerge under complete information. If there is no uncertainty, then either the status quo is stable, or it is not. Bluffing and related strategems are logically precluded, and outright conflict never occurs. Precisely because information is complete, a challenger will not rationally take an aggressive action unless it is prepared to endure conflict, and a defender will not be able to manipulate a challenger's behavior when its willingness, or unwillingness, to retaliate is common knowledge.

---

[6]For some recent attempts to unravel the empirical correlates of these puzzles, see Betts (1987); Geller (1990); George and Smoke (1974); Huth (1988a, 1988b, 1990); Huth and Russett (1984, 1988); Kugler (1984); Organski and Kugler (1980); Weede (1981, 1983); and Wu (1990). For a discussion of the theoretical relevance of this literature, see Kilgour and Zagare (1992b).

[7]For the interested reader, we summarize the intermediate case of one-sided incomplete information in footnote 10.

[8]To simplify the representation, we assume that Defender prefers CD to CC.

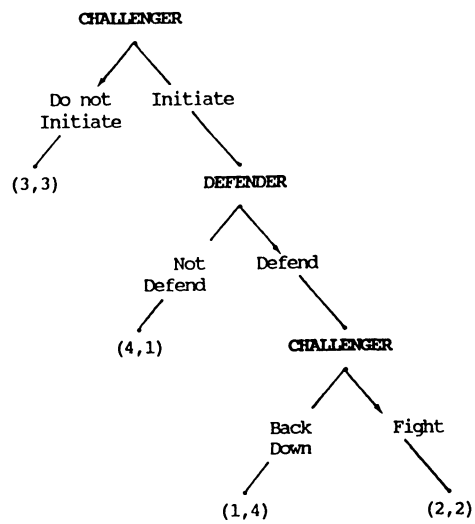
[9]Similar conclusions obtain for mutual (symmetric) deterrence games (Zagare, 1987).

TABLE 1. Logically possible outcome orderings of asymmetric deterrence games with complete information.

	Ch	*Def*	*Ch*	*Def*	*Ch*	*Def*	*Ch*	*Def*
p_{Ch} and p_{Def}	1	1	1	0	0	1	0	0
Possible outcomes								
CC	3	4 or 3	3	4 or 3	3	4 or 3	3	4 or 3
DD	2	2	2	1	1	2	1	1
CD	1	3 or 4	1	3 or 4	2	3 or 4	2	3 or 4
DC	4	1	4	2	4	1	4	2
Subgame perfect equilibrium	CC		DC		CC		CC	
Predicted outcome	stable deterrence		Challenger wins		stable deterrence		stable deterrence	

Key: Ch = Challenger
Def = Defender
p_{Ch} = probability Challenger's threat is credible
p_{Def} = probability Defender's threat is credible

4 = best, 3 = next-best, 2 = next-worst, 1 = worst



Key: (x, y) = (Challenger, Defender)
4 = best, 3 = next-best, 2 = next-worst, 1 = worst
→ = rational choice

FIG. 3. Asymmetric deterrence with credible threats and complete information.

Incomplete Information and Deterrence Stability

So far we have shown that the credibility of the players' threats completely determines the outcome of an asymmetric deterrence game of complete information. But what if the players do not have certain knowledge of each other's preference for conflict or capitulation? Under these conditions, a player's rational behavior must depend upon two kinds of information: its own preferences, and its *beliefs* about its opponent's preferences, that is, its opponent's credibility.

Recall that we distinguish two *types* of players: a **Hard** player prefers conflict to capitulation; a **Soft** player does not. Each player knows its own type, but not its opponent's. We assume that the players have made probability estimates about these critical distinctions: Challenger (respectively, Defender) (initially) is believed to be Hard with probability p_{Ch} [p_{Def}], and Soft otherwise. Thus, p_{Ch} and p_{Def} describe the extent to which the players are perceived to be Hard. The higher these quantities, the more credible the players' threat is. Over the full range of these perceptions, what should rational players do?

Before answering this question fully, note that a player's rational actions are partly determined by the player's own preferences. For example, at the last node of the tree, a Hard Challenger will always Fight to achieve $C_2 = c_2^+$ rather than 0; analogously, a Soft Challenger will always Back Down to achieve 0 rather than $C_2 = c_2^-$.

Similarly, if there is a challenge, a Hard Defender will always defend because both possible outcomes associated with this choice, CD and DD, with utilities 1 and $D_2 = d_2^+$, are preferred by a Hard Defender to capitulation, DC, which has utility 0. This will not be true, however, if Defender is Soft. A Soft Defender prefers CD to DC to DD, so its choice will rationally depend on its estimate of how likely it is to receive 1 (at CD) or $D_2 = d_2^-$ (at DD) if it defends. This estimate depends, in turn, on its estimate of Challenger's type, which, as just shown, directly determines the final outcome when Defender resists. At the second node of the tree, therefore, a Hard Defender always defends, but a Soft Defender's rational choice might be either to defend or to not defend.

Given that Challenger initiates, what information should Defender use to estimate the probability that the Challenger is Hard? It is rational that Defender update its initial estimate, p_{Ch}, on the basis of the new information it has received. Specifically, Defender now knows that Challenger has Initiated, and this may rationally change Defender's beliefs about Challenger's type.

Finally, what choice should Challenger make at the top of the tree? A Challenger of either type could rationally initiate a challenge, or not. This decision depends upon its estimates of *both* Defender's type *and* its anticipated behavior. With probability p_{Def}, Defender is Hard and will defend with certainty. But because the choice of a Soft Defender is uncertain (see above), Challenger must estimate the probability that Defender is Soft but will defend nonetheless.

In summary, rational choices in an asymmetric deterrence game of incomplete information are determined by the players' preferences for conflict (DD) versus capitulation (either CD or DC), and also by each player's probability estimate, updated when appropriate, of the credibility of the other's threat.

Rational Deterrence

Up to this point we have shown the implications of threat credibility for deterrence in a game of complete information. We have also discussed what rational players would need to consider when information, and hence credibility, is uncertain. Now we determine fully the rational choices for each player in an asymmetric deterrence game of incomplete information.

The results are shown in Figure 4, which represents p_{Ch}, the probability that Challenger is Hard, along the horizontal axis, and p_{Def}, the probability that Defender is Hard, along the vertical axis. On these two axes are indicated several constants, such as d*, which are defined and fully discussed in the Appendix. These constants are convenient thresholds for categorizing and interpreting the equilibria of asymmetric deterrence games of incomplete information.[10]

As noted above, p_{Def} and p_{Ch} can range from 0 to 1, as the corresponding threats range from perfectly incredible to perfectly credible. The four corners of this figure thus correspond to the four games of complete information discussed previously. Specifically, the northeast (respectively, northwest, southeast, and southwest) corner represents the situation where each player (only Defender, only Challenger, and neither player) has a credible threat. Recall that the status quo is stable in three of these corners, but not in the southeast corner where Challenger's threat alone is credible.

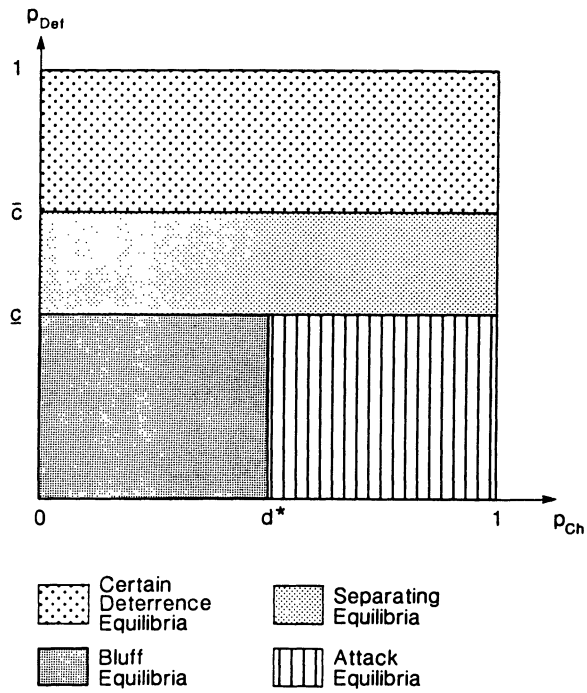


FIG. 4. Location of equilibria.

[10]In fact, two of the three thresholds are characteristic of one-sided games of incomplete information, as follows:

a. When Defender's preferences are common knowledge, but not Challenger's, the status quo remains stable (*i. e.*, deterrence works), if Defender is Hard, or if Defender is Soft and Challenger's threat is not credible enough to dissuade Defender from resisting ($p_{Ch} < d^*$). But when Defender is Soft and the Challenger's threat is sufficiently credible (*i.e.*, $p_{Ch} > d^*$), Challenger gains the advantage and the outcome is DC.

b. When only Challenger's preferences are common knowledge, deterrence succeeds when Challenger is Soft, or when Defender's threat is credible enough ($p_{Def} > \bar{c}$) to deter a Challenger of any type. But if Challenger is Hard and Defender's credibility falls below this threshold (i.e., $p_{Def} < \bar{c}$), then Challenger wins (CD) when Defender is Soft and conflict (DD) results when Defender is Hard.

It is proven in the Appendix that four major types of (perfect Bayesian) equilibria are possible for intermediate values of threat credibility.[11] These equilibria are represented by probability combinations (x_H, x_S, y_S, p) where

x_H = the probability that a Hard Challenger will Initiate
x_S = the probability that a Soft Challenger will Initiate
y_S = the probability that a Soft Defender will Defend
p = Defender's conditional probability that Challenger is Hard, given that Challenger has initiated

The first three probabilities are strategic variables describing Challenger's and Defender's choices. The fourth probability is the *a posteriori* assessment of Challenger's credibility, updated by Defender after Challenger has been observed to Initiate. Equilibrium values, or ranges, for p are reported in the Appendix, but are indicated in the text only when relevant to the discussion. Because these are perfect Bayesian equilibria, each player always acts so as to maximize its expected utility given its current beliefs about the type of opponent, and updates these beliefs rationally (*i.e.*, uses Bayesian updating), based on its observations. Thus, the equilibria encompass all possible patterns of behavior consistent with basic rationality assumptions. Through the equilibria one can determine the logical implications of the players' beliefs. We describe next the four major equilibrium categories, the conditions under which they arise, and the consequences for deterrence in each case.

Deterrence Equilibria

Deterrence equilibria are equilibria in which Challenger never initiates a crisis, that is, $x_H = x_S = 0$. In other words, at a deterrence equilibrium, only the status quo outcome can result. Challenger's decision is independent of its type, but it may be contingent on Defender's strategy, that is, on y_S. (See the Appendix for details.) Although the status quo may sometimes result from other equilibria, all other equilibria carry the possibility of other outcomes, depending on the players' types, beliefs, and choices. Thus, the status quo is fully robust under a deterrence equilibrium, and not under other kinds.

A deterrence equilibrium is a plausible descriptive of the relationship between hostile powers during periods of relative détente. Under such circumstances, any challenge to the status quo would at best gain little, and would risk a great deal. On the other hand, the status quo might still survive under less benevolent conditions such as those associated with the most heated periods of the Cold War. A superpower might have been deterred from directly challenging the other, but only after seriously considering other courses of action. Similarly, a deterrence equilibrium could be associated with strategic relationships characterized by clear-cut military superiority—assuming that the Defender is the more powerful player—whereas a more tenuous status quo outcome might persist for a time in a contentious parity relationship such as that of Great Britain and Germany early in the twentieth century.

Interestingly, deterrence equilibria can come into play under *any* set of player beliefs about threat credibility, save for the pure case (discussed above) in which Challenger's threat is perfectly credible, and Defender's is not. This means that deterrence could conceivably emerge under (almost) any conditions in an

[11]There is another equilibrium type, but, as explained in the Appendix, it almost never occurs.

asymmetric relationship. As demonstrated in the Appendix, the key to this stability may be the willingness of a Soft Defender to defend with a high enough probability. (Recall that a Hard Defender always defends.) To support this willingness, Defender must believe, even after a crisis has been initiated, that Challenger is unlikely to be Hard.

Although deterrence is (almost) always possible, it is not so often inevitable. In this regard, we distinguish two types of deterrence equilibria: **Certain** and **Steadfast**. We next discuss their characteristics.

Certain Deterrence. When Defender's credibility, p_{Def}, is high enough, the only equilibrium of the asymmetric deterrence game is a deterrence equilibrium. We call this equilibrium Certain Deterrence; when it exists, the status quo is the only rational outcome of the game. The reason is simple: when a Certain Deterrence Equilibrium exists, there are no other equilibria, so there are no other rational behavior possibilities.

Certain Deterrence Equilibria ($x_H = 0$, $x_S = 0$, y_S unrestricted) occupy the entire upper region of the unit square in Figure 4. Notice that the existence of a Certain Deterrence Equilibrium does not depend on the *a priori* credibility of Challenger's threat. In other words, this equilibrium is invariant with respect to Challenger's credibility: when a Certain Deterrence Equilibrium occurs, no rational Challenger—Hard or Soft—will choose to Initiate.

Because the existence of a Certain Deterrence Equilibrium does not depend on the Challenger's credibility, it should not be surprising that Defender's credibility is critical. As shown in the Appendix, for a Certain Deterrence Equilibrium to exist,

$$p_{Def} > \bar{c} = \frac{1 - c_3}{1 - c_2^+}.$$

In other words, Defender's *a priori* credibility must exceed this threshold. The level of this threshold, and hence the likelihood of certain deterrence, depends only on Challenger's evaluation of the status quo (c_3) and the value of a Hard Challenger for conflict (c_2^+), relative to victory (1) and capitulation (0). Note that as the value of the status quo increases, the minimum for p_{Def} decreases. This means that the more a challenger values the status quo, the larger the region and hence the more likely deterrence will be the only possible outcome. In particular, a defender can make deterrence more likely either by raising the credibility of its threat past this threshold (with, for instance, a public commitment to defend), or, in what Snyder and Diesing (1977:ch. 3) would call an "accommodative move," by enhancing the challenger's evaluation of the status quo by granting trade or other concessions.

It is perhaps unfortunate that for most of the Cold War period decision-makers in both the Soviet Union and the United States concentrated on the former method and ignored the latter in molding the strategic relationship of the superpowers. Although manipulation of the status quo is not generally thought of as a strategem for practitioners of deterrence, our model supports George and Smoke's (1974:531) observation that efforts to this effect will reduce the need for overt deterrent threats and increase the likelihood that these more traditional deterrence tactics will succeed when and if they are practiced. In this context it is important to note that moves that increase a challenger's satisfaction with the status quo do not necessarily come at a defender's expense. Mutually agreed upon arms reductions, for instance, might benefit both parties simultaneously, thereby making deterrence more likely.

The prospects of Certain Deterrence are also enhanced by decreasing the value of c_2^+ and, significantly, not decreasing the value of c_2^-. This means that deterrence

becomes more likely as the value a Hard Challenger associates with conflict approaches the value it associates with capitulation. In other words, when conflict offers minimal advantages, deterrence becomes more certain. This seems incontrovertible enough.

Nevertheless, indiscriminate increases in the cost of conflict do not necessarily contribute to the likelihood of deterrence. Thus, like George and Smoke (1974:507), we find that deterrent "threats are often irrelevant or dysfunctional." When the challenger is Soft, and already prefers capitulation to conflict, further increases in the cost of conflict (*i.e.*, the value c_2-) are redundant and unnecessary. This suggests that when other tactics (see above) are unavailable, a prudent defender has no reason not to pursue a policy of minimum deterrence, which relies "on the retention of only enough nuclear weapons to provide an assured destruction capability" (Kegley and Wittkopf, 1989:351), rather than a maximum deterrent strategy rooted in an overkill capability. Thus, the normative implications of our model—rooted in beliefs and perceptions about the nature of deterrent threats—are at odds with those of models such as Intriligator and Brito's (1981, 1984), which ignore this variable and examine the strategic consequences associated with different cost levels of conflict. In such models, increases in the cost of conflict beyond the point where players would ever choose to fight produce monotonic increases in the likelihood of stable deterrence (Kugler and Zagare, 1987, 1990).

It is worth pointing out that the threshold that defines the region of Certain Deterrence is analogous to the two thresholds that define what we call **Sure-thing** deterrence in a mutual deterrence model in which either player can challenge the other initially (Kilgour and Zagare, 1991). Of course, in this symmetric model, either player can take on the role of Defender.

This is *not* to say that the challenger's credibility is of little moment in an asymmetric deterrence situation. As we show below, Challenger, too, can potentially influence the outcome of an asymmetric deterrence game by manipulating Defender's perception of the likelihood it is Hard. But when Defender's credibility is sufficiently high, deterrence stability is absolute, and never hinges on Challenger's willingness to fight.

One interesting question our model highlights but does not answer is the connection between a player's offensive capability and its defensive credibility. When these variables are independent, there is no trade-off between increasing the costs of conflict (up to a point) and increasing the perception of credibility to augment stability. Such will be the case when the primary costs of warfare are inflicted by one's opponent.

On the other hand, a defender faces a dilemma when these costs are primarily public, and borne by the players more or less equally. In this case, attempts to increase the costs of conflict will also lead to a loss of credibility and, ultimately, to less-probable deterrence. Again, this suggests that an overkill capability—when coupled with shared-cost, consequential damage (such as damage to the biosphere)—is not conducive to deterrence stability (Anderton and Fogarty, 1990).

Finally, note that the existence of a Certain Deterrence Equilibrium does not depend upon a particular behavior plan that a Soft Defender may have (*i.e.*, on y_S). Intuitively, the reason is that Certain Deterrence occurs only when Challenger's probability that Defender is Hard—and therefore will defend its interests—is high. Consequently, Challenger assigns very little weight to the benefits it might receive should it Initiate against a Soft Defender. However, the weight assigned these benefits may be crucial under other conditions.

Steadfast Deterrence. A deterrence equilibrium may still exist even when Defender's credibility falls below the threshold required for Certain Deterrence.

But then, a deterrence equilibrium cannot occur alone—it co-exists with equilibria of other types (see below), so its occurrence in actual play is far from certain. Moreover, for this type of deterrence equilibrium to occur, Defender must be steadfast in the sense of being committed to defend with at least a threshold probability, even when it is Soft.

A Steadfast Deterrence Equilibrium [$x_H = 0$, $x_S = 0$, $y_S \geq y^*(p_{Def})$] may come into play even when a challenger places a relatively low value on the status quo, or a relatively high value on conflict. This form of deterrence occurs only when Defender's threat is less credible than required for a Certain Deterrence Equilibrium. The difference is that with Steadfast Deterrence Defender's threat to defend when Hard is not in itself sufficient to sustain the status quo. Further commitment is necessary. Specifically, to offset the relative decline in Defender's credibility (*i.e.*, the probability that it is Hard), Challenger must believe there is a high enough probability (y_S) that a Soft Defender will resist. To rationally support this intention, Defender must believe that it is fairly likely that any Challenger who initiates is Soft and will back down if Defender defends. Thus, the *a posteriori* credibility of **Challenger's** retaliatory threat is critical. (See the Appendix for details.)

Under the proper conditions, a Steadfast Deterrence Equilibrium might evolve quite naturally. Suppose, for instance, that Challenger is not satisfied and is considering a challenge. Defender may likely be Soft, yet manage to convince Challenger that it doubts Challenger's resolve and will therefore defend with a high enough probability, even if Soft. Faced with a high probability of resistance, Challenger now finds its second-best outcome, the status quo, very appealing.

By their very nature, actual examples of deterrence equilibria (Certain or Steadfast) are difficult to identify.[12] Nevertheless, one indication that a Steadfast Deterrence Equilibrium is in play, or that a defender is trying to induce one, is a public denigration of the capability and, ultimately, the credibility of a challenger's threat. For example, in the 1950s, when Mao—on more than one occasion—expressed open and unequivocal doubts about U.S. resolve, he was probably trying to deter a possible coercive move by the United States. From the Chinese point of view, it may have been strategically immaterial whether the United States was, or was not, a "paper tiger," or whether U.S. decision-makers thought the Chinese were Soft. What was important to the Chinese was to convince U.S. leaders that they thought the United States was very likely Soft. Similarly, early in the postwar period, Soviet declaratory policy denying the strategic significance of nuclear weapons may have been consistent with actual Soviet beliefs. But if not, strategic considerations probably dictated its content.

Other Equilibria

In addition to deterrence equilibria, three other kinds of equilibria may occur in an asymmetric deterrence game: Separating, Bluff, and Attack. We now consider their logical and strategic properties serially.[13]

Separating Equilibria. Separating Equilibria ($x_H = 1$, $x_S = 0$, $y_S = 0$) lie between Certain Deterrence Equilibria and Attack and Bluff Equilibria (see Figure 4). At a

[12]Achen and Snidal (1989:161) give as possibilities "the first Soviet–American War which erupted over Hungary in 1956. . . . The second one (over Chile) in the early 1970s. . . . The U.S.–China War, which began when the United States bombed the North Vietnamese dikes . . . [and] the second Korean War." For a debate on this issue, see Huth and Russett (1990) and Lebow and Stein (1990).

[13]We ignore here the Transitional Equilibria (discussed in the Appendix) that occur only on the boundary between Certain Deterrence and Separating and Attack Equilibria.

Separating Equilibrium, the players' preferences are fully revealed by their strategy choices: a Hard Challenger always initiates and a Soft Challenger never does. Likewise, if challenged, a Hard Defender always resists and a Soft Defender always capitulates. The status quo may remain stable under a Separating Equilibrium, but only when Challenger is Soft.[14] When Challenger is Hard, it gains an advantage if Defender is Soft, and precipitates a conflict when Defender is Hard. Thus, three of the four possible outcomes of an asymmetric deterrence game (see Figure 2) can transpire under a Separating Equilibrium.

Separating Equilibria separate players by type. Under Separating Equilibria, more so than any other, the stability of the status quo depends on Challenger's actual preferences. This suggests that when Separating Equilibria exist, attempts by Defender to manipulate Challenger's type can be expected. For example, a defender might try to influence the domestic political process of a potential challenger, as the Vietminh did in late 1953 in order to bring the French to the negotiation table at Geneva (Zagare, 1979). But the converse is also true. When Challenger is Hard, the final outcome of the game will be determined by Defender's type. The challenger will then have an interest in promoting soft-liners in Defender's bureaucracy, because Soft defenders will appease a challenger whereas Hard defenders will willingly endure conflict (Snyder and Diesing, 1977).

It is noteworthy that Separating Equilibria are the only equilibria in incomplete information deterrence games that do not correspond to one of the four games of complete information discussed above. One of the contributions of this analysis, therefore, is to uncover this qualitatively different strategic environment, and the prototypical strategies associated with it.

When the status quo is stable, Separating Equilibria may be indistinguishable from Deterrence and Bluff Equilibria (see below). Separating Equilibria are most likely to be observed, however, when a challenger's regime and policy orientation (and little else) change simultaneously. For instance, in 1953, the temporary shift away from the confrontational policies of Stalin's Soviet Union and toward those of the new collective leadership which sought accommodation with the West in Vietnam and elsewhere suggests that a Separating Equilibrium may have been in play. Similarly, during the 1967 crisis in the Middle East, Israel's policies shifted dramatically from submission to confrontation when Moshe Dayan, a well-known hard-liner, replaced Prime Minister Levi Eshkol as Defense Minister (Zagare, 1981). This sudden change in Israel's strategy choice is consistent with the existence of a Separating Equilibrium.

Separating Equilibria occur in an intermediate range of Defender's credibility, not high enough to ensure deterrence but not low enough to make the preservation of the status quo virtually impossible. The upper bound of the region of Separating Equilibria ($\bar{c}$) coincides with the lower bound of the Certain Deterrence region. The lower bound of the Separating region is the threshold:

$$p_{\text{Def}} \geq \underline{c} = 1 - c_3 .$$

Thus, the initial credibility requirements of Defender's threat are directly related to the Challenger's evaluation of the status quo, c_3. As this value increases, approaching that of Challenger's best outcome, the lower bound of the region of Separating Equilibria moves downward, shrinking the regions of Attack and Bluff Equilibria where conflict is more likely (see below).

[14]Hence the survival of the status quo does not imply that a Deterrence Equilibrium is in play.

Bluff Equilibria. When Defender's credibility is below $\underline{c}$, the lower boundary for Separating Equilibria, then, depending upon how credible *Challenger's* threat is, one of two new equilibria arises: Bluff or Attack.

As Figure 4 reveals, Bluff Equilibria [$x_H = 1$, $x_S = x$ (p_{Ch}), $y_S = y$ (p_{Def})] occur when both Defender's and Challenger's credibility is relatively low, that is, when both players believe that the other probably prefers to capitulate rather than fight. As we have seen, in the extreme (complete information) case when each player's threat is simply not credible, the status quo is stable and Challenger is deterred. Such is not necessarily the case, however, when the players are uncertain of each other's preferences.

At a Bluff Equilibrium, players' behaviors depend on their types. The Challenger initiates for certain in the unlikely event that it is Hard. (After all, chances are that Defender is Soft and likely to capitulate.) But if Challenger is Soft, it adopts a mixed strategy, initiating with some positive probability. The more credible it is perceived to be, the greater this probability.

The equilibrium choice of a Hard Defender is, as always, to defend. But at a Bluff Equilibrium, even a Soft Defender defends with some positive probability. This *conditional* probability, y_S, is a function of Defender's initial credibility, just as Challenger's conditional probability, x_S, is a function of Challenger's credibility. But the lower Defender's credibility, the greater its tendency to bluff and resist a challenge when Soft!

In fact, the family of Bluff Equilibria is *pooling* for Defender—the *unconditional* (*i. e.*, without regard to Defender's type) *a priori* probability that Defender chooses to defend does not depend on its credibility (see Appendix). In other words, at any Bluff Equilibrium, Defender's overall probability of capitulating or resisting if challenged is always the same, regardless of the value of p_{Def}. The Bluff Equilibria do not present this property from Challenger's point of view—in fact, the overall probability of a challenge is directly proportional to the Challenger's credibility, p_{Ch}.

The logic of this equilibrium configuration guarantees that Challenger always faces the same probability that Defender will defend. This strategem serves to conceal Defender's type if it is Soft—it is not possible to infer that a Defender who resists is likely or unlikely to be Hard. As a consequence, Challenger is less and less willing to risk a challenge as its own credibility decreases—after all, the probability that it will have to back down if it challenges becomes greater and greater as its credibility drops. But the only way that Defender can achieve this constant level of defense readiness is to be prepared to defend when Soft more frequently the lower its credibility.

More than at any other equilibrium, then, play under a Bluff Equilibrium is likely to be a "competition in risk-taking" (Schelling, 1960, 1966). In one sense, this should be no surprise since the complete information analogue of such games bears a structural resemblance to the game of Chicken. Nonetheless, our model provides additional insight into the precise set of conditions under which "manipulative bargaining tactics" (Young, 1975) are relevant to the behavior of states in an acute crisis (Snyder and Diesing, 1977).

In contrast to the conventional wisdom, however, our model reveals an advantage for Defender since Defender's type and, ultimately, its choice, is critical in determining the actual outcome. Of course, if Challenger does not initiate, a crisis never occurs; and if it does, and Defender does not defend, the bluff will have succeeded and Challenger will gain an advantage. But if Defender defends, a rational Soft Challenger will back down, so that Defender will win. In neither case, though, will conflict result unless Challenger is prepared for it. Thus, given the postulated sequence of choices, even if Challenger initiates, it will not necessarily win. This may be one reason why "commitment" and related bluffing tactics,

although seductive, have rarely been used by challengers in precipitating a crisis (Snyder and Diesing, 1977:227–231) and why, at least in the Nuclear Age, decision-makers have sought "to retain wide freedom of choice as long as possible and to avoid becoming boxed into an irrevocable position" (Young, 1968:218; see also George and Smoke, 1974:531).

With the benefit of hindsight, it is plausible to associate many of the events punctuating the U.S.–Soviet relationship during the 1950s and 1960s with bluff conditions. Starting with the Berlin crisis of 1948, the Soviets and the Chinese precipitated a number of confrontations designed to probe the limits of U.S. resolve. When the United States stood firm, they backed down. Although one cannot say for sure just what the actual American preferences were, the challengers' preference for capitulation over conflict was revealed by their choices. In these cases at least, they were simply bluffing (Betts, 1987:108).

Attack Equilibria. Like Bluff Equilibria, Attack Equilibria ($x_H = 1$, $x_S = 1$, $y_S = 0$) occur only when the credibility of Defender is low. What distinguishes the two equilibria is the perceived credibility of Challenger. When Challenger's credibility is relatively low—like Defender's—a Bluff Equilibrium arises; when it exceeds a certain threshold, the Attack Equilibrium choices predominate.

At an Attack Equilibrium, Challenger—whatever its type—always initiates, and a Soft Defender always capitulates. Thus, because a Hard Defender always defends, war will occur if and only if a Hard Challenger challenges a Hard Defender. Whereas this is unlikely under attack conditions, actual conflict remains possible. Typically, a defender has few options and little defense. Like the Americans during the crises in Hungary in 1956 and in Czechoslovakia in 1968, and like the Soviets during the 1956 Suez crisis, a defender can only accept the inevitable; any other reaction would be contrary to its interests.

When an Attack Equilibrium is in play, the status quo is never stable. Defender would prefer any other equilibrium to an Attack Equilibrium; in fact, Defender's expected utility is usually less at this equilibrium than at any other. The Attack Equilibrium is the only equilibrium where there is no chance that Challenger will accept the status quo.

There are essentially two ways Defender can shift the game away from an Attack Equilibrium. By increasing its own credibility or Challenger's evaluation of the status quo, it could induce a Separating Equilibrium. But no matter what equilibrium is in play, these tactics never disadvantage a defender. Alternatively, a defender afraid of an Attack Equilibrium evolving could try to induce a Bluff Equilibrium. As Figure 4 indicates, if Challenger's *a priori* credibility, p_{Ch}, falls below

$$d^* = \frac{1}{1 - d_2^-}$$

there will be a Bluff Equilibrium; otherwise an Attack Equilibrium will come into play (see Appendix for details). This means that an Attack Equilibrium is more likely the more costly conflict is to a Soft Defender. In other words, the higher the costs of confrontation, the more likely an Attack Equilibrium; the lower, the more likely a Bluff Equilibrium.

This observation is consistent with the argument of some strategic thinkers that one consequence of nuclear weapons has been to make conflicts in areas peripheral to a defender's interests more likely. In this sense, at least, nuclear weapons are destabilizing. For instance, the Soviets might have been less willing to invade Afghanistan in 1979 if the world were not nuclear, simply because the United States would have been more likely, *ceteris paribus,* to offer resistance were it not facing a nuclear power. Much the same could be said about the American

involvement in Vietnam. Thus, our model explains why nuclear weapons may contribute, simultaneously, to the stability of "basic," "passive," or "Type I" deterrence, and to the instability of "extended," "active," or "Type II and Type III" deterrence (Kahn, 1960, 1965; Betts, 1987).

Multiple Equilibria

As indicated above, a Steadfast Deterrence Equilibrium exists simultaneously only with a Separating, an Attack, or a Bluff Equilibrium. Consequently, the question arises as to which equilibrium is likely to come into play, and when. We address this question now.

When a Separating Equilibrium co-exists with a Steadfast Deterrence Equilibrium, the deterrence equilibrium is unlikely as long as Challenger is Hard. This is because a Hard Challenger strictly prefers a Separating Equilibrium, and a Soft Challenger is indifferent. Moreover, the Hard Challenger can choose to initiate (for certain), so the status quo is unlikely to persist under rational play. (Of course, the actual outcome depends upon both Challenger's and Defender's actions.)

This reasoning can be formalized, using refinements of Nash equilibria (van Damme, 1991; Fudenberg and Tirole, 1991), by formulating the asymmetric deterrence game of incomplete information as a signaling game in which Challenger may initiate to signal its type to a Soft Defender. (At equilibrium, a Hard Defender always resists, so it is known to be unreceptive to signals.) When a Separating Equilibrium exists, Steadfast Deterrence fails the so-called Intuitive Criterion, because only a Hard Challenger could gain by signaling. This means that, in the presence of the Separating Equilibrium, Steadfast Deterrence is inconsistent with more refined notions of rational behavior; in this sense, our unique prediction is for a Separating Equilibrium within the central band of Figure 4.

This use of Nash equilibrium refinements does not, however, rule out Steadfast Deterrence in comparison with the Bluff or Attack Equilibria. In this case, both types of Challengers could gain from signaling, but if it sustains a great enough belief that any challenge is from a Soft Challenger, a Soft Defender could rationally resist. In this circumstance, it appears that there are no equilibrium refinements that rule out either Steadfast Deterrence or the Bluff or Attack Equilibrium.

When a Steadfast Deterrence Equilibrium and a Bluff Equilibrium co-exist, can they be distinguished by behavior? The answer is certainly yes if Challenger is Hard, for then it will initiate a crisis for certain. Although a Soft Challenger is indifferent as to the two equilibria, its corresponding strategies—and their associated outcomes—are different. Specifically, if Challenger selects a strategy consistent with deterrence, it never challenges, and the status quo remains stable. But if a Soft Challenger selects a strategy consistent with a Bluff equilibrium, it initiates with some positive probability (depending on how credible its threat is). Thus, when Challenger is Soft, the status quo is much less likely to remain stable at a Bluff Equilibrium. Whereas this choice may well result in Defender's capitulation, a confrontation will surely follow when Defender is actually Hard. Nonetheless, given Challenger's indifference, it may be difficult to say which equilibrium is being played.

Such is not the case, however, when the Attack Equilibrium and the Steadfast Deterrence Equilibrium co-exist. Under the Attack Equilibrium, nothing like deterrence can ever occur. A Challenger of either type will initiate a confrontation, so the status quo will be overturned with certainty.

Finally, it should be noted that there is another feature of the Steadfast Deterrence Equilibrium that may help to identify it. This equilibrium depends on beliefs concerning events off the equilibrium path—specifically, Defender must have a high conditional probability that any Challenger who did initiate would be

Soft, although in fact no Challenger ever initiates. It is these beliefs that prepare even a Soft Defender to respond to any challenge. Defender's high level of confidence that it will face down its adversary may be a distinguishing feature of Steadfast Deterrence.

A possible indicator that a Defender in fact possesses such beliefs is an intelligence failure associated with a surprise attack. For instance, despite strong signals to the contrary, American decision-makers failed to give credence to the possibility that Japan would attack in 1941 or that North Korea would invade the South in 1950; moreover, both American and Israeli intelligence were caught off guard by the Arab attack in 1973. It seems plausible to suggest that in each of these cases the Defender believed that a Steadfast Equilibrium was in play.

Summary and Conclusions

This paper explores the connection between threat credibility and deterrence stability in simple asymmetric deterrence games. These games model the choices of a Challenger and a Defender to foment or avoid a crisis. Each player is either Hard (preferring confrontation to capitulation) or Soft (preferring the opposite). Under complete information, Hard players have completely credible (*i.e.*, rational) deterrent threats and Soft players have completely incredible threats. But under incomplete information, a player knows its own preferences but is uncertain of its adversary's, so that threat credibility depends upon each player's perception of the likelihood that the other is Hard. The higher the perceived probability that a player would actually prefer to execute its deterrent threat, the higher its credibility, and conversely.

In this model, Challenger may choose either to accept the status quo or to initiate a crisis; in the latter case, Defender may capitulate or defend; if it defends, Challenger either backs down or fights. Unlike other models of crisis initiation, then, the model we develop makes no fixed behavioral assumption about the strategy choices of the players at *any* stage of the game; nor does it place *a priori* restrictions on critical preference relationships. This permits us to explore the full range of potential crisis situations, to offer a more general assessment of the conditions under which the status quo is likely to persist, and to present a more complete description of the circumstances and consequences of a deterrence failure.

One important contribution of this work is the precision it adds to the debate about the stability of asymmetric deterrence relationships, a precision that one strategic analyst recently pointed out is "extraordinarily difficult" to come by (Betts, 1987:ix). This precision goes beyond merely identifying certain critical thresholds that define the various equilibrium regions and their associated behavioral patterns. Our simple model allows us also to specify the interplay of the players' belief systems, their evaluation of the various outcomes (including the status quo), and the credibility of their threats in determining the dynamics of an asymmetric deterrent relationship.

In this context it is worth noting that the conclusions we draw are at once readily interpretable and, for the most part, unambiguous. This is, perhaps, still another sense in which our model adds precision to the analysis of asymmetric deterrence games. In order to penetrate the underlying dynamic of an important category of deterrence relationships, our plan has been to focus on the simplest possible case, and resist unduly complicating the model. We hope that what has been sacrificed in the way of complexity has been more than offset by economy of expression and by the clarity of our conclusions. These conclusions flow from the identification and categorization of all perfect Bayesian equilibria of the simple deterrence game we postulate.

We found that deterrence is possible under almost all conditions explored by our model, although some conditions are much more likely to support a stable status quo than others. In general, deterrence becomes more probable as the Challenger's evaluation of the status quo increases, as its perception of the Defender's credibility grows, and as the benefits of conflict decline.

Stable deterrence is one consequence of a Certain Deterrence Equilibrium. Certain Deterrence Equilibria are the only equilibria that do not depend upon Challenger's type, Defender's commitments, or specific beliefs of the Defender about the Challenger's willingness to fight rather than back down. The only critical variable here, it turns out, is the initial credibility of Defender's threat. When it is high enough, a deterrence outcome is certain and the survival of the status quo is assured. Significantly, our model reveals that increasing the costs of conflict does not necessarily lead to increases in strategic stability. Past a certain point, such increases are either unnecessary or counterproductive, leading us to recommend minimum deterrent policies that are effective and costly enough to ensure deterrence, yet not so incredible as to undermine it.

At Separating Equilibria, deterrence fails with certainty when Challenger is Hard, but occurs with certainty when it is Soft. The stability of the status quo, therefore, directly and immediately depends on Challenger's type. A Defender's incentive to promote the political standing of doves among Challenger's decision-making elite (Snyder and Diesing, 1977:297–310) is likely strongest under conditions associated with the existence of Separating Equilibria.

Separating Equilibria exist at an intermediate level of Defender credibility, high enough to make a stable status quo feasible, but not high enough to make it certain. Of the four equilibrium categories we identify, only Separating Equilibria do not correspond to one of four possible asymmetric games of complete information. Thus, our model has uncovered a new form of rational strategic behavior that might arise in an asymmetric deterrence situation.

At lower levels of Defender credibility, two additional equilibrium categories exist: Bluff and Attack. At an Attack Equilibrium, Challenger always initiates a crisis and Defender backs down if Soft and defends if Hard. At a Bluff Equilibrium, a Hard Challenger initiates with certainty, whereas a Soft Challenger initiates sometimes in the hope that Defender will back down. In turn, a Hard Defender always defends and a Soft Defender defends only occasionally. We identify a Bluff Equilibrium as a "competition in risk-taking," and relate the levels of those risks to the players' initial credibilities. Thus, deterrence—defined in terms of the survival of the status quo—is most problematic when either of these equilibria come into play.

One difference between Bluff and Attack situations is the level of Challenger credibility associated with each. Attack, and the certainty of a challenge, is associated with high Challenger credibility. Bluff occurs under the opposite conditions. In our model, the threshold credibility level varies with the costs of conflict to a Soft Defender (relative to capitulation and the status quo). The higher these costs, the more likely that Challenger will initiate conflict. This, we argue, accounts for the large number of crises and brush-fire and proxy wars in the nuclear period. Thus, whereas nuclear weapons may enhance central deterrence under certain conditions, they may also have the opposite effect in areas peripheral to Great-Power rivalries.

Appendix

In this Appendix, we analyze the game described in the text. For convenience, a complete formal description of the game will be given first.

The two players of the game are called *Ch* (Challenger) and *Def* (Defender). In the first stage of the game, *Ch* chooses whether to Initiate *(I)* or not $(\bar{I})$. Given *I* is chosen, *Def* then decides whether to Defend *(D)* or not $(\bar{D})$. Given *D* is chosen, *Ch* must then elect to Back Down *(B)* or to Fight *(F)*. The game ends as soon as $\bar{I}$, $\bar{D}$, *B*, or *F* is selected; the corresponding utilities to *(Ch, Def)* are (c_3, d_3), (1, 0), (0, 1), and (C_2, D_2), respectively. The game tree is shown in Figure A1.

The two types of each player are called Hard *(H)* and Soft *(S)*. Player *Ch* is Hard with probability p_{Ch} and Soft with probability $1 - p_{Ch}$; for Player *Def*, the probabilities are p_{Def} and $1 - p_{Def}$, respectively. The types of the players determine the values of the random variables C_2 and D_2. Specifically,

$$C_2 = \begin{cases} c_2^+ \text{ with probability } p_{Ch} \\ c_2^- \text{ with probability } 1 - p_{Ch} \end{cases}$$

and, independently,

$$D_2 = \begin{cases} d_2^+ \text{ with probability } p_{Def} \\ d_2^- \text{ with probability } 1 - p_{Def} \end{cases}.$$

The parameters of the game satisfy

$$c_2^- < 0 < c_2^+ < c_3 < 1;$$
$$d_2^- < 0 < d_2^+ < \min\{d_3, 1\};$$
$$0 \le p_{Ch} \le 1;\ 0 \le p_{Def} \le 1.$$

As is usual in games of incomplete information, each player learns its own type prior to the start of play, but is ignorant of the type of the opponent.

Our objective is to identify and study all perfect Bayesian equilibria of this game. Under this equilibrium definition (see Tirole [1988] or Rasmusen [1989] for

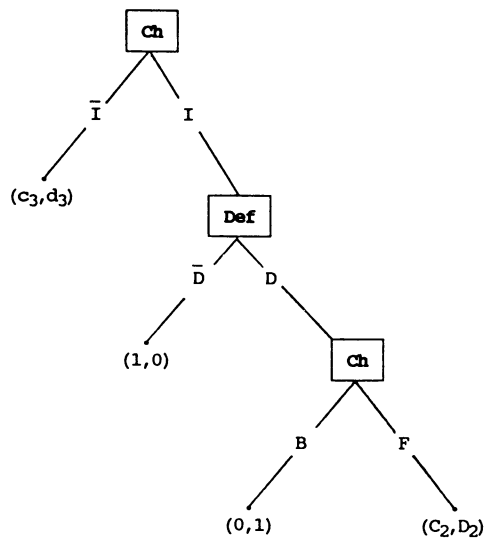


FIG. A1. Asymmetric deterrence game.

details), each player's action at any time during the game is always chosen to maximize expected utility, which is evaluated using Bayesian updating of the player's belief about the opponent's type. In other words, a player's belief about the type of the opponent reflects not only the initial probabilities but also the actions that the opponent has already taken.

Begin at the lowest decision node (*Ch* chooses *F* or *B*) in the tree of Figure A1. Because *Ch* receives C_2 for *F* or 0 for *B*, it is clear that *Ch* chooses *F* if *Ch* is Hard because $C_2 = c_2^+ > 0$, and *Ch* chooses *B* if *Ch* is Soft because $C_2 = c_2^- < 0$.

Now consider the second decision node (*Def* chooses *D* or $\bar{D}$) in Figure A1. If *Def* chooses $\bar{D}$ it receives 0, whereas choosing *D* results in D_2 if *Ch* is Hard and 1 if *Ch* is Soft. If *Def* is Hard, $D_2 = d_2^+ > 0$, and it is clear that *Def* always does better with *D* than with $\bar{D}$. But if *Def* is Soft, its best choice is not determined by dominance; in this case, denote the probability that *Def* chooses *D* by the strategic variable y_S.

At the top decision node in Figure A1 (*Ch* chooses *I* or $\bar{I}$), *Ch*'s decision may depend on its type. Denote the probability that *Ch* chooses *I* by x_H if *Ch* is Hard, and x_S if *Ch* is Soft. It is easy to calculate that *Ch*'s expected utility at this node is

$$E_{Ch|H} = (1 - x_H)(c_3) + x_H[(1 - p_{Def})(1 - y_S)(1) + (p_{Def} + y_S - p_{Def}y_S)(c_2^+)] \quad \text{(A1)}$$
$$E_{Ch|S} = (1 - x_S)(c_3) + x_S[(1 - p_{Def})(1 - y_S)(1) + (p_{Def} + y_S - p_{Def}y_S)(0)] \quad \text{(A2)}$$

if *Ch* is Hard and Soft, respectively. Player *Ch*'s optimal choices of x_H and x_S, to maximize $E_{Ch|H}$ and $E_{Ch|S}$, will be considered below.

Now consider again *Def*'s choice at the second node, in the case where *Def* is Soft. Let *p* denote *Def*'s conditional probability that *Ch* is Hard, given that the second node has been reached. Because *Def*'s initial probability on this event must have been p_{Ch}, Bayesian updating requires that

$$\text{if } x_H + x_S > 0,\ p = \frac{p_{Ch}x_H}{p_{Ch}x_H + (1 - p_{Ch})x_S}. \quad \text{(A3)}$$

Note that if $x_H + x_S = 0$, then both types of *Ch* choose $\bar{I}$ for certain, and the second node is off the equilibrium path. In this case, condition (A3) is void.

At the second node, a Soft *Def*'s expected utility is then

$$E_{Def|S} = y_S[p \bullet d_2^- + (1 - p) \bullet 1] + (1 - y_S) \bullet 0 = y_S[pd_2^- + (1 - p)].$$

The perfectness requirement then places the following restrictions on y_S at any equilibrium:

$$\begin{aligned} &\text{if } y_S = 0,\ p \geq d^*, \\ &\text{if } y_S = 1,\ p \leq d^*, \\ &\text{if } 0 < y_S < 1,\ p = d^* \end{aligned} \quad \text{(A4)}$$

where $d^* = 1/(1 - d_2^-)$.

We now return to the analysis of *Ch*'s decision at the top node of Figure A1. From (A1), it follows that

$$E_{Ch|H} = (1 - x_H)c_3 + x_H[c_2^+ + (1 - p_{Def})(1 - y_S)(1 - c_2^+)]$$

so that

$$\frac{\partial E_{Ch|H}}{\partial x_H} = -c_3 + c_2^+ + (1 - p_{Def})(1 - y_S)(1 - c_2^+).$$

It follows that, at equilibrium,

$$\begin{aligned} &\text{if } x_H = 1,\ (1 - p_{Def})(1 - y_S) \geq c^*, \\ &\text{if } x_H = 0,\ (1 - p_{Def})(1 - y_S) \leq c^*, \\ &\text{if } 0 < x_H < 1, (1 - p_{Def})(1 - y_S) = c^*, \end{aligned} \tag{A5}$$

where $c^* = \dfrac{c_3 - c_2^+}{1 - c_2^+}$. Analogous consideration of (A2) for a Soft *Ch* yields the equilibrium conditions

$$\begin{aligned} &\text{if } x_S = 1,\ (1 - p_{Def})(1 - y_S) \geq c_3, \\ &\text{if } x_S = 0,\ (1 - p_{Def})(1 - y_S) \leq c_3, \\ &\text{if } 0 < x_S < 1,\ (1 - p_{Def})(1 - y_S) = c_3. \end{aligned} \tag{A6}$$

A perfect Bayesian equilibrium of the game can be represented by a set of probabilities (x_H, x_S, y_S, p) satisfying (A3)–(A6). We will find all probability combinations obeying these conditions. In every case, it will be easy to verify that a perfect Bayesian equilibrium has been attained. Of course, to determine the actual play of the game at any such equilibrium, the observations above must be applied—a Hard *Def* always Defends, a Hard *Ch* always Fights, and a Soft *Ch* always Backs Down. We now undertake a systematic search for all perfect Bayesian equilibria.

The following lemma will make the search for equilibria easier:

Lemma: Suppose x_H and x_S are probabilities satisfying (A5) and (A6). Then
(i) if $x_S > 0$, $x_H = 1$;
(ii) if $x_H < 1$, $x_S = 0$.
Proof: Both (i) and (ii) follow immediately from the fact that $c^* < c_3$, which is easy to verify directly.

The equilibria are most conveniently classified using the values of x_H and x_S; the lemma therefore reduces the number of classes.

Deterrence Equilibria ($x_H = x_S = 0$)

Any equilibrium with $x_H = x_S = 0$ is called a Deterrence Equilibrium, as it always ends in the status quo. For such an equilibrium, the Bayesian updating condition (A3) does not apply because the second decision node (Figure A1) is off the equilibrium path.

For a Deterrence Equilibrium with $y_S = 1$, (A4) requires that $p \leq d^*$. Because $(1 - p_{Def})(1 - y_S) = 0 < c^* < c_3$, (A5) and (A6) are satisfied automatically. For a Deterrence Equilibrium with $y_S = 0$, (A4) implies $p \geq d^*$, and (A5) is satisfied iff (if and only if) $p_{Def} \geq \dfrac{1 - c_3}{1 - c_2^+} = \bar{c}$. Of course, if (A5) holds, (A6) does also. Finally, for a Deterrence Equilibrium with $0 < y_S < 1$, conditions (A4) and (A5) yield $p = d^*$ and $y_S \geq y^*(p_{Def}) = 1 - \dfrac{c^*}{1 - p_{Def}}$. It is easy to verify that $y^*(0) = 1 - c^* = \bar{c}$, $y^*(p)$ is a decreasing function of p, and $y^*(\bar{c}) = 0$.

In summary, there is a *Deterrence Equilibrium* ($x_H = 0$, $x_S = 0$) for any values of p_{Ch} and p_{Def}. If $p_{Def} \geq \bar{c}$, the Deterrence Equilibria are described by $y_S = 0$, $p \geq d^*$; $0 < y_S < 1$, $p = d^*$; and $y_S = 1$, $p \leq d^*$. If $p_{Def} < \bar{c}$, the Deterrence Equilibria are given by $y^*(p_{Def}) \leq y_S < 1$, $p = d^*$; and $y_S = 1$, $p \leq d^*$.

Separating Equilibrium ($x_H = 1$, $x_S = 0$)

From (A3), $p = 1$ at any Separating Equilibrium, so $y_S = 0$ is the only possibility consistent with (A4). But now (A5) shows that $x_H = 1$ can occur only if $1 - p_{Def} \geq c^*$, which is equivalent to $p_{Def} \leq \bar{c}$. Similarly, (A6) shows that $x_S = 0$ can occur only if $1 - p_{Def} \leq c_3$, which is equivalent to $p_{Def} \geq \underline{c} = 1 - c_3$. Note that $0 < \underline{c} < \bar{c} < 1$.

In summary, there is a *Separating Equilibrium* ($x_H = 1$, $x_S = 0$) iff $\underline{c} \leq p_{Def} \leq \bar{c}$. In this case the unique Separating Equilibrium has $y_S = 0$ and $p = 1$.

Attack Equilibrium ($x_H = x_S = 1$)

Condition (A3) implies that $p = p_{Ch}$ at any Attack Equilibrium. It is clear that $y_S \neq 1$, because $y_S = 1$ is inconsistent with $x_S = 1$, by (A6). Suppose that $y_S = 0$. Then $p_{Ch} \geq d^*$, by (A4). Also $x_S = 1$ only if $1 - p_{Def} \geq c_3$, or $p_{Def} \leq \underline{c}$, by (A6). The lemma shows that $x_H = 1$ if $x_S = 1$, so (A5) need not be considered.

Now suppose that $0 < y_S < 1$. Then (A4) shows that $p_{Ch} = d^*$, and (A6) permits $x_S = 1$ iff $(1 - p_{Def})(1 - y_S) \geq c_3$, which is equivalent to $y_S \leq y(p_{Def}) = 1 - \dfrac{c_3}{1 - p_{Def}}$. Clearly the latter condition can be met iff $y(p_{Def}) > 0$, which is true iff $p_{Def} < 1 - c_3 = \underline{c}$.

In summary, there is an *Attack Equilibrium* ($x_H = x_S = 1$) iff $p_{Ch} \geq d^*$ and $p_{Def} \leq \underline{c}$. Either $y_S = 0$ and $p = p_{Ch}$ or if $p_{Ch} = d^*$ and $p_{Ch} < \underline{c}$, $0 < y_S \leq y(p_{Def})$ and $p = d^*$.

Transitional Equilibrium ($0 < x_H < 1$, $x_S = 0$)

By (A3), $p = 1$ at any Transitional Equilibrium, so $y_S = 0$ by (A4). The lemma and (A5) show that the only remaining condition is $1 - p_{Def} = c^*$, which is equivalent to $p_{Def} = \bar{c}$.

In summary, there is a *Transitional Equilibrium* ($0 < x_H < 1$, $x_S = 0$) iff $p_{Def} = \bar{c}$. In this case, the value of x_H is unrestricted, but $y_S = 0$ and $p = 1$.

Bluff Equilibrium ($x_H = 1$, $0 < x_S < 1$)

Note first that, at a Bluff Equilibrium, $p = \dfrac{p_{Ch}}{p_{Ch} + (1 - p_{Ch})x_S}$, by (A3). Now (A6) shows that $(1 - p_{Ch})(1 - y_S) = c_3$ is required for $0 < x_S < 1$; this condition is equivalent to $y_S = y(p_{Def}) = 1 - \dfrac{c_3}{1 - p_{Def}}$. Because $y_S \geq 0$, $1 - \dfrac{c_3}{1 - p_{Def}} \leq 1$, which is equivalent to $p_{Def} \leq \underline{c}$, is necessary. By the lemma, the only remaining condition to be applied is (A4).

Because $y(p_{Def}) < 1$, $y_S = 1$ is impossible. In order that $y_S = y(p_{Def}) = 0$, $p_{Def} = \underline{c}$ is required. Furthermore, (A4) implies that $p \geq d^* = \dfrac{1}{1 - d_2}$, which is equivalent to $x_S \leq x(p_{Ch}) = -\left(\dfrac{p_{Ch}}{1 - p_{Ch}}\right) d_2$. Similarly, for $0 < y_S < 1$, $p = d^*$ and $p_{Def} < \bar{c}$ are required. Therefore, $x_S = x(p_{Ch})$ from (A4). Note that $x(p_{Ch}) \geq 0$, $x(p_{Ch}) < 1$ iff $p_{Ch} < d^*$, and $x(p_{Ch}) = 1$ if $p_{Ch} = d^*$.

TABLE A1.

Name	*Existence conditions*	*Strategies*	$E_{Ch\mid S}$	$E_{Ch\mid H}$	$E_{Def\mid S}$	$E_{Def\mid H}$
(Certain) Deterrence (Steadfast)	$p_{Def} \geq \overline{c}$	$x_H = 0,\ x_S = 0$ y_S unrestricted	c_3	c_3	d_3	d_3
	$P_{Def} < \overline{c}$	$x_H = 0,\ x_S = 0$ $y_S \geq y^*(p_{Def})$				
Separating	$\underline{c} \leq p_{Def} \leq \overline{c}$	$x_H = 1$ $x_S = 0$ $y_S = 0$	c_3	$1 - p_{Def}(1 - c_2^+)$	$(1 - p_{Ch})\, d_3$	$(1 - p_{Ch})\, d_3 + p_{Ch} d_2^+$
Attack	$p_{Def} \leq \underline{c}$ and $p_{Ch} \geq d^*$	$x_H = 1$ $x_S = 1$ $y_S = 0$	$1 - p_{Def}$	$1 - p_{Def}(1 - c_2^+)$	0	$(1 - p_{Ch}) + p_{Ch} d_2^+$
Bluff	$p_{Def} \leq \underline{c}$ and $p_{Ch} \leq d^*$	$x_H = 1$ $x_S = x(p_{Ch})$ $y_S = y(p_{Def})$	c_3	$c_3 + (1 - c_3)\, c_2^+$	$d_3\,(1 - p_{Ch} + p_{Ch} d_2^-)$	$d_3(1 - p_{Ch} + p_{Ch} d_2^-) + p_{Ch}(d_2^+ - d_2^-)$

In summary, there is a *Bluff Equilibrium* ($x_H = 1, 0 < x_S < 1$) iff $p_{Def} \leq \underline{c}$ and if $p_{Def} < \underline{c}$, $p_{Ch} \leq d^*$. If $p_{Def} = \underline{c}$, these equilibria satisfy $y_S = 0$, $p \geq d^*$, and $x_S \leq x(p_{Ch}) = -\left(\frac{p_{Ch}}{1 - p_{Ch}}\right) d_2^-$. (The latter condition actually restricts x_S iff $p_{Ch} < d^*$.) If $p_{Def} < \underline{c}$, the Bluff Equilibria satisfy $y_S = y(p_{Def})$, $p = d^*$, and $x_S = x(p_{Ch})$. (The latter group of equilibria exist only when $x(p_{Ch}) < 1$, or $p_{Ch} < d^*$.)

Parenthetically, it is worth noting that, where $p_{Def} < \underline{c}$ and $p_{Ch} < d^*$, the Bluff Equilibrium is pooling for Defender, in that the *a priori* probability that Defender will defend is $p_{Def} \bullet 1 + (1 - p_{Def}) \bullet y(p_{Def}) = \overline{c}$, independent of p_{Def}. But the *a priori* probability of initiation is $p_{Ch} \bullet 1 + (1 - p_{Ch}) \bullet x(p_{Ch}) = p_{Ch}(1 - d_2^-)$, so an analogous property does not hold for the Challenger. In fact, the frequency of initiation is proportional to Challenger's credibility in this region.

This completes the determination of all of the perfect Bayesian equilibria of the game defined in the text. The equilibria that exist on sets of positive measure in the (p_{Ch}, p_{Def})-square are indicated in Figure 4 of the text. Note that the only Deterrence Equilibria shown in Figure 4 are those with $y_S = 0$. In fact, a Deterrence Equilibrium can always be constructed at any point of the (p_{Ch}, p_{Def})-square by making p small enough and y_S large enough. These equilibria are summarized and the corresponding payoffs—to both types of each player—are shown in Table A1.

We now consider the players' preferences about equilibria that co-exist. Such co-existence can occur in essentially one way—if $p_{Def} < \overline{c}$, there is always a Deterrence Equilibrium along with one other equilibrium which may be Separating, Attack, or Bluff. (In this discussion, sets of measure zero in the (p_{Ch}, p_{Def})-square will be ignored.) Using Table A1, it is possible to assess the players' relative preferences between the competing equilibria as a function of their types.

If $\underline{c} \leq p_{Def} \leq \overline{c}$, a Deterrence Equilibrium and a Separating Equilibrium co-exist. In this case, a Soft *Ch* is indifferent between the two equilibria, a Hard *Ch* prefers the Separating Equilibrium, and *Def*, whether Hard or Soft, prefers the Deterrence Equilibrium.

If $p_{Def} \leq \underline{c}$ under $p_{Ch} \geq d^*$, a Deterrence Equilibrium and an Attack Equilibrium co-exist. In this case, both types of *Ch* strictly prefer the Attack Equilibrium, and a Soft *Def* strictly prefers the Deterrence Equilibrium. A Hard *Def* prefers the Deterrence Equilibrium if p_{Ch} is high enough $\left(p_{Ch} \geq \frac{1 - d_3}{1 - d_2^-}\right)$, but may in fact prefer the Attack Equilibrium if p_{Ch} is relatively low.

Finally, a Deterrence Equilibrium coincides with a Bluff Equilibrium if $p_{Def} \leq \underline{c}$ and $p_{Ch} \leq d^*$. A Soft *Ch* is indifferent between the two equilibria, whereas a Hard *Ch* strictly prefers the Bluff Equilibrium. A Soft *Def* prefers the Deterrence Equilibrium, as does a Hard *Def* who has a relatively large d_3. But if d_3 is small enough, a Hard *Def* prefers the Bluff Equilibrium.

References

ACHEN, C. H., AND D. SNIDAL (1989) Rational Deterrence Theory and Comparative Case Studies. *World Politics* **41**:143–169.

ANDERTON, C. H., AND T. FOGARTY (1990) Consequential Damage and Nuclear Deterrence. *Conflict Management and Peace Science* **11**:1–15.

BANKS, J. S. (1990) Equilibrium Behavior in Crisis Bargaining Games. *American Journal of Political Science* **34**:599–614.

BETTS, R. K. (1987) *Nuclear Blackmail and Nuclear Balance.* Washington, DC: Brookings Institution.

BUENO DE MESQUITA, B., AND D. LALMAN (1992) *War and Reason: A Confrontation between Domestic and International Imperatives.* New Haven, CT: Yale University Press.

BUENO DE MESQUITA, B., AND J. D. MORROW (1991) Capabilities, Perception and Escalation: Testing Limited-Information Hypotheses about Crises. Unpublished.

DORAN, C. F. (1989a) "Power Cycle Theory of System Structure and Stability: Commonalities and Complementarities." In *Handbook of War Studies,* edited by M. I. Midlarsky. Boston: Unwin Hyman.

DORAN, C. F. (1989b) Systemic Disequilibrium, Foreign Policy Role, and the Power Cycle. *Journal of Conflict Resolution* **33**:371–401.

ELLSBERG, D. (1959) The Theory and Practice of Blackmail. Lecture at the Lowell Institute, Boston, MA, March 10. Reprinted in *Bargaining: Formal Theories of Negotiation,* edited by O. R. Young. Urbana: University of Illinois Press, 1975.

FEARON, J. D. (1990) Deterrence and the Spiral Model: The Role of Costly Signals in Crisis Bargaining. Paper presented at the Annual Meeting of the American Political Science Association.

FUDENBERG, D., AND J. TIROLE (1991) *Game Theory.* Cambridge, MA: MIT Press.

GELLER, D. S. (1990) Nuclear Weapons, Deterrence, and Crisis Escalation. *Journal of Conflict Resolution* **34**:291–310.

GEORGE, A. L., AND R. SMOKE (1974) *Deterrence in American Foreign Policy.* New York: Columbia University Press.

GILPIN, R. (1975) *U.S. Power and the Multilateral Corporation.* New York: Basic Books.

GILPIN, R. (1981) *War and Change in World Politics.* New York: Cambridge University Press.

HERMANN, C. F. (1969) "International Crisis as a Situational Variable." In *International Politics and Foreign Policy,* edited by J. N. Rosenau. New York: Free Press.

HUTH, P. (1988a) *Extended Deterrence and the Prevention of War.* New Haven, CT: Yale University Press.

HUTH, P. (1988b) Extended Deterrence and the Outbreak of War. *American Political Science Review* **82**:423–443.

HUTH, P. (1990) The Extended Deterrent Value of Nuclear Weapons. *Journal of Conflict Resolution* **34**:270–290.

HUTH, P., AND B. RUSSETT (1984) What Makes Deterrence Work. *World Politics* **36**:496–526.

HUTH, P., AND B. RUSSETT (1988) Deterrence Failure and Crisis Escalation. *International Studies Quarterly* **32**:29–45.

HUTH, P., AND B. RUSSETT (1990) Testing Deterrence Theory: Rigor Makes a Difference. *World Politics* **42**:466–501.

INTRILIGATOR, M. D., AND D. L. BRITO (1981) Nuclear Proliferation and the Probability of Nuclear War. *Public Choice* **37**:247–260.

INTRILIGATOR, M. D., AND D. L. BRITO (1984) Can Arms Races Lead to the Outbreak of War? *Journal of Conflict Resolution* **28**:63–84.

KAHN, H. (1960) *On Thermonuclear War.* Princeton, NJ: Princeton University Press.

KAHN, H. (1965) *On Escalation: Metaphors and Scenarios.* Rev. ed. Baltimore, MD: Penguin Books.

KEGLEY, C. W., AND E. WITTKOPF (1989) *The Nuclear Reader: Strategy, Weapons, War.* 2d ed. New York: St. Martin's Press.

KILGOUR, D. M., AND F. C. ZAGARE (1991) Uncertainty and Deterrence. *American Journal of Political Science* **35**:303–334.

KILGOUR, D. M., AND F. C. ZAGARE (1992a) Modeling "Massive Retaliation." Unpublished.

KILGOUR, D. M., AND F. C. ZAGARE (1992b) Uncertainty and the Role of the Pawn in Extended Deterrence. Unpublished.

KIM, W., and J. D. MORROW (1990) When Do Power Transitions Lead to War? Paper presented at the Annual Meeting of the Midwest Political Science Association.

KINDLEBERGER, C. P. (1974) *The World in Depression 1929–1939.* Berkeley: University of California Press.

KINDLEBERGER, C. P. (1976) "Systems of International Economic Organization." In *Money and the Coming World Order,* edited by D. P. Calleo. New York: New York University Press.

KISSINGER, H. (1992) U.S. Needs Europe to Avoid Becoming Second-Class Power. *Buffalo News,* March 8:G-9.

KRASNER, S. D. (1976) State Power and the Structure of International Trade. *World Politics* **28**:317–347.

KUGLER, J. (1984) Terror without Deterrence. *Journal of Conflict Resolution* **28**:470–506.

KUGLER, J., AND F. C. ZAGARE (EDS.) (1987) *Exploring the Stability of Deterrence.* University of Denver Graduate School of International Studies Monograph Series in World Affairs. Boulder, CO: Lynne Rienner.

KUGLER, J., AND F. C. ZAGARE (1990) The Long-Term Stability of Deterrence. *International Interactions* **15**:255–278.

LEBOW, R. N. (1981) *Between Peace and War: The Nature of International Crisis.* Baltimore, MD: Johns Hopkins University Press.

LEBOW, R. N., AND J. G. STEIN (1990) Deterrence: The Elusive Dependent Variable. *World Politics* **42**:336–369.

LEVY, J. S. (1987) Declining Power and the Preventive Motivation for War. *World Politics* **40**:82–107.

LUCE, R. D., AND H. RAIFFA (1957) *Games and Decisions: Introduction and Critical Survey.* New York: Wiley.

MEARSHEIMER, J. J. (1990) Back to the Future: Instability in Europe after the Cold War. *International Security* **15**:5–56.

MIDLARSKY, M. I. (1988) *The Onset of War.* Boston: Unwin Hyman.

MODELSKI, G. (1983) "Long Cycles of World Leadership." In *Contending Approaches to World Systems Analysis,* edited by W. R. Thompson. Beverly Hills, CA: Sage.

MODELSKI, G., AND W. R. THOMPSON (1989) "Long Cycles and Global War." In *Handbook of War Studies,* edited by M. I. Midlarsky. Boston: Unwin Hyman.

MORGAN, P. M. (1977) *Deterrence: A Conceptual Analysis.* Beverly Hills, CA: Sage.

MORROW, J. D. (1989a) Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining. *American Journal of Political Science* **33**:941–972.

MORROW, J. D. (1989b) "Bargaining in Repeated Crises: A Limited Information Model." In *Models of Strategic Choice in Politics,* edited by P. C. Ordeshook. Ann Arbor: University of Michigan Press.

NALEBUFF, B. (1986) Brinkmanship and Nuclear Deterrence: The Neutrality of Escalation. *Conflict Management and Peace Science* **9**:19–30.

ORDESHOOK, P. C. (ED.) (1989) *Models of Strategic Choice in Politics.* Ann Arbor: University of Michigan Press.

ORGANSKI, A. F. K., AND J. KUGLER (1980) *The War Ledger.* Chicago: University of Chicago Press.

POWELL, R. (1990) *Nuclear Deterrence Theory: The Search for Credibility.* New York: Cambridge University Press.

RASMUSEN, E. (1989) *Games and Information.* New York: Basil Blackwell.

SCHELLING, T. C. (1960) *The Strategy of Conflict.* Cambridge, MA: Harvard University Press.

SCHELLING, T. C. (1966) *Arms and Influence.* New Haven, CT: Yale University Press.

SMOKE, R. (1987) *National Security and the Nuclear Dilemma.* Reading, MA: Addison-Wesley.

SNYDER, G. H., AND P. DIESING (1977) *Conflict among Nations: Bargaining, Decision Making and System Structure in International Crises.* Princeton, NJ: Princeton University Press.

TIROLE, J. (1988) *The Theory of Industrial Organization.* Cambridge, MA: MIT Press.

VAN DAMME, E. (1991) Refinements of Nash Equilibria. No. 9107, Center for Economic Research, Tilberg University, Netherlands.

WAGNER, R. H. (1991) Nuclear Deterrence, Counterforce Strategies, and the Incentive to Strike First. *American Political Science Review* **83**:727–749.

WALTZ, K. N. (1979) *Theory of International Politics.* Reading, MA: Addison-Wesley.

WEEDE, E. (1981) Preventing War by Nuclear Deterrence or by Detente. *Conflict Management and Peace Science* **6**:1–8.

WEEDE, E. (1983) Extended Deterrence by Superpower Alliance. *Journal of Conflict Resolution* **27**:231–253.

WU, S. S. G. (1990) To Attack or Not to Attack: A Theory and Empirical Assessment of Extended Immediate Deterrence. *Journal of Conflict Resolution* **34**:531–552.

YOUNG, O. R. (1968) *The Politics of Force: Bargaining during International Crises.* Princeton, NJ: Princeton University Press.

YOUNG, O. R. (1975) (ED.) *Bargaining: Formal Theories of Negotiation.* Urbana: University of Illinois Press.

ZAGARE, F. C. (1979) The Geneva Conference of 1954: A Case of Tacit Deception. *International Studies Quarterly* **23**:390–411.

ZAGARE, F. C. (1981) Nonmyopic Equilibria and the Middle East Crisis of 1967. *Conflict Management and Peace Science* **5(Spring)**:139–162.

ZAGARE, F. C. (1987) *The Dynamics of Deterrence.* Chicago: University of Chicago Press.

ZAGARE, F. C. (1990) Rationality and Deterrence. *World Politics* **42**:238–260.

ZAGARE, F. C. (1992) The Rites of Passage: Parity, Nuclear Deterrence, and Power Transitions. State University of New York at Buffalo. Typescript.