JOURNAL OF THEORETICAL POLITICS

VOLUME 10 NUMBER 1 JANUARY 1998

SAGE Publications
ISSN 0951-6928

Journal of Theoretical Politics 10(1): 59–87
0951–6928[1998/01]10:1; 59–87; 002150

DETERRENCE THEORY AND THE SPIRAL MODEL REVISITED

Frank C. Zagare and D. Marc Kilgour

ABSTRACT

The theoretical literature of interstate conflict is dominated by two conceptual models, classical deterrence theory and the spiral model. The fundamental tenet of classical deterrence theory is that credible and capable threats can prevent the initiation, and contain the escalation, of conflict. By contrast, proponents of the spiral model claim that the prescriptions associated with deterrence theory frequently lead to vicious cycles of reciprocated conflict.

According to Jervis 'both sets of theorists fail to discuss the conditions under which their theories will not apply'. In this article we do just that, identifying and comparing the conditions associated with conflict spirals and with crisis stability, in the context of a game-theoretic escalation model with incomplete information. For the special case in which a challenger is likely willing to endure an all-out conflict, our analysis indicates that the conditions associated with successful deterrence, limited conflict, and escalated conflict are mutually exclusive.

KEY WORDS • deterrence theory • escalation • game theory • limited wars • spiral model

The theoretical literature of interstate conflict is dominated by two conceptual models, classical deterrence theory and the spiral model. A fundamental tenet of *classical deterrence theory* is that credible and capable threats can prevent the initiation, and contain the escalation, of conflict (Zagare, 1996). For example, deterrence theorists argue that the Second World War might have been prevented if Britain had clearly communicated its intention to resist German expansionism.

By contrast, proponents of the *spiral model* claim that the prescriptions associated with deterrence theory frequently lead to vicious cycles of reciprocated conflict.[1] Conflicts like the First World War spiral out of

This material is based upon work supported by the National Science Foundation under Grant No. SBR-9514160 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 and the Social Sciences and Humanities Research Council of Canada.

1. Like classical deterrence theory (Zagare, 1996), the spiral model is difficult to define precisely. Jervis (1976) understandably uses broad brush strokes in his discussion. Our understanding is equally broad, encompassing reaction-process models of both arms races and of conflict escalation. The common thread is that of an unintended, semi-automatic, intensification of conflict.

control when states inadvertently threaten each other's security while communicating deterrent threats or when acting to shore up their credibility. Thus, an enemy is provoked and deterrence fails.

According to Jervis (1976: 95) 'both sets of theorists fail to discuss the conditions under which their theories will not apply'. In this paper we do just that, identifying and comparing the conditions associated with conflict spirals and with crisis stability in the context of a game-theoretic extended deterrence/escalation model with incomplete information. For the special case in which a challenger is likely willing to endure an all-out conflict, our analysis indicates that the conditions associated with successful deterrence, limited conflict, and escalated conflict are mutually exclusive. Our analysis not only provides insight into the dynamics of interstate conflict, but also subsumes within a single theoretical framework two conceptual models long thought incompatible.

1. The Two-level Asymmetric Escalation Game

In an anarchical self-help system in which each state 'must rely on [its] own strength and art for caution against all others' (Hobbes, [1651] 1968: 224), a countervailing threat is frequently prescribed as the most effective stratagem for maintaining order. Only the likelihood of a strong response, it is argued, can deter attacks and stabilize the status quo. From Sun Tzu and Vegetius on to Brodie and his intellectual descendants, statesmen have been counseled that to bring about peace, one must prepare for war. Those who fail to heed the *para bellum* dictum are considered in dereliction of duty.

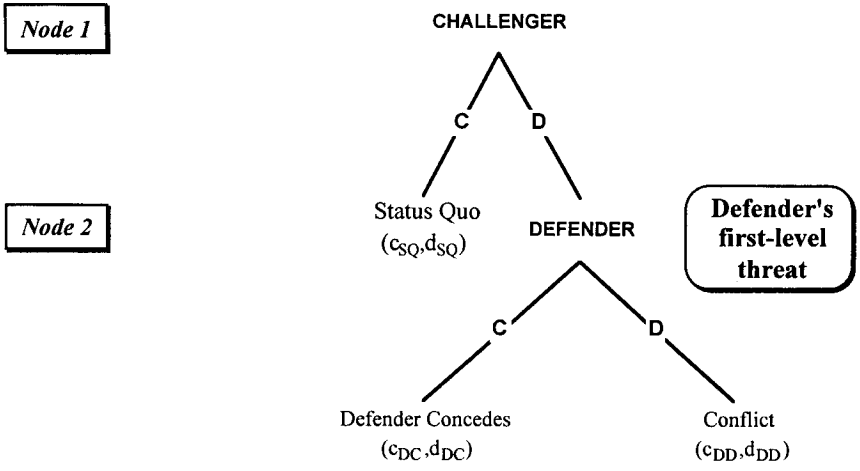
Of course, being willing and able to resist attacks does not always prevent them. There is no magic elixir here. Arms races frequently spiral, conflicts intensify, crises escalate, and wars erupt. Why? Jervis (1978) and numerous others argue that conflict is inherent in the interstate system. In seeking security, states face a troubling dilemma: even entirely defensive measures – with no aim other than self-preservation – may set off a sequence of actions and reactions that ends only in war. In other words, when a state moves to fend off a potential attack, it may inadvertently threaten the security interests of another, leading to a response, then a counter-response, a counter-counter-response, and so on, until an all-out conflict occurs.

When do deterrent threats work? Under what conditions can conflict be managed, and when does it inevitably spiral out of control? Our purpose is to answer these questions by identifying the conditions associated with successful deterrence, limited conflicts, and escalation spirals. To this end we explore a game model of extended deterrence and escalation called the

'Asymmetric Escalation Game'. This model, represented in Figure 3, was developed and described in detail in Zagare and Kilgour (1996).

We have chosen to work within a model of extended deterrence and escalation for one simple reason: with the exception of the Franco–Prussian war, all nine major power wars since the Congress of Vienna have involved an extended deterrence failure (Danilovic, 1995). Thus, classical deterrence theory seems most likely to fail, and the spiral model most likely to apply, in the context of extended deterrence.

To describe the Asymmetric Escalation Game model, and to motivate our analysis, we begin with Figure 1, a model of a rather primitive form of deterrence, and a plausible representation of the Truman administration's interpretation of the US–USSR strategic relationship in 1947. In this rudimentary model, there are two players, Challenger and Defender. Challenger begins play at Node 1 either by cooperating (C) and accepting



Key: C = Cooperate/Concede; D = Demand/Defy

Assumptions: Challenger: DC > SQ (Definition of Challenger)
SQ > DD (Defender's *first-level* threat is capable)
Defender: SQ > DC and DD (Definition of Defender)

Defender's types: 1. Hard: DD > DC (first-level threat is credible)
2. Soft: DC > DD (first-level threat lacks credibility)

Figure 1. Containment circa 1947

the Status Quo (outcome SQ) or by defecting (D) and demanding a change in the existing order.[2]

If Challenger cooperates, the game ends and the payoffs to Challenger and Defender are c_{SQ} and d_{SQ} respectively. (In general, the players' utilities at outcome K will be denoted (c_K, d_K).) But if Challenger defects, Defender must decide how to react. At Node 2 Defender can either concede (C) to Challenger's demand or defy (D) Challenger. Concession leads to outcome DC (Defender Concedes) while defiance results in outcome DD (Conflict or, more mnemonically, Defender Defies). Defender's choice at Node 2 constitutes its first-level threat. Notice that there are no escalation choices in this simple model and, hence, no possibility of a conflict spiral.

To transform this simple model into a deterrence game, we make three assumptions about the players' preferences. First, we assume Challenger prefers Defender Concedes to the Status Quo, i.e. prefers DC to SQ. This restriction on preferences provides Challenger with an immediate incentive to defect. Second, we assume Defender prefers the Status Quo to all other outcomes, i.e. prefers SQ to DC and DD. This assumption, in effect, makes deterrence Defender's principal objective. Finally, we assume that Challenger prefers the Status Quo (SQ) to Conflict (DD).

The last assumption is an important one that deserves special comment. Challenger's presumed preference for the Status Quo over Conflict, perforce, affords Defender a first-level threat that is capable, i.e., a threat that, if carried out, hurts Challenger (Schelling, 1966). We assume capability for a very straightforward reason: without it, deterrence cannot possibly succeed (Zagare, 1987: Ch. 4).[3] We believe that most empirical exceptions to conclusions we draw – both here and elsewhere – about the likelihood of limited conflicts occur when this significant limiting condition is not satisfied (Zagare and Kilgour, 1995: 400–1).

No fixed assumption is made, however, about Defender's preference between Defender Concedes (DC) and Conflict (DD). This, *the* critical preference relationship of the model, is allowed to vary. The relationship is critical because it establishes the actual *credibility* of Defender's first-level threat (i.e. Defender's *type* – Hard or Soft) which, under complete

2. Note that while Challenger has no direct attack option in this model, an initial assault on an ally or client state (i.e. a pawn) could be construed as a demand for change. In our opinion, most games in which a frontal attack is a viable option do not involve extended deterrence. Direct deterrence relationships in which a frontal attack is a real possibility are modeled in Kilgour and Zagare (1991) and Zagare and Kilgour (1993a).

3. In other words, a capable threat is a necessary condition for successful deterrence.

information, determines the outcome of the game.[4] Logically, there are only two possibilities:

1. Defender prefers DD to DC. In this case it is rational for Defender to carry out its threat should Challenger defect at Node 1. Threats that are rational to carry out are called *credible* (Selten, 1975; Kilgour and Zagare, 1991). Players who prefer to carry out a threat are called *Hard*. In the game of Figure 1, deterrence succeeds (i.e. the outcome is SQ) when Defender is known to be Hard.[5]
2. Defender prefers DC to DD. In this case, Defender's threat lacks credibility since Defender would prefer *not* to carry it out. Players who prefer not to execute a threat are called *Soft*. When Defender is known to be Soft, deterrence fails and the outcome is DC.

Given the above it is easy to understand why, during the early years of the Cold War, containment of the Soviet Union was seen as a straightforward engineering problem. After all, until 1949 the United States had a monopoly on atomic weapons and was clearly the world's dominant industrial and political power. The credibility of the US threat to defend itself, or its most important allies, was taken to be almost self-evident; but even when it was not, US credibility could easily be shored up by words, or deeds, or both.

For example, in 1947 when Greece and Turkey were thought to be threatened, the Truman Doctrine was proclaimed: military and economic help was to be provided to any country resisting outside (i.e. 'communist') aggression. During the Berlin crisis of 1948, a more forceful message was sent when the United States transferred several B-29s, the so-called atomic bombers, to British and German bases. The intent of this signal was obvious: the US was Hard.

For a while, containment worked, or at least it seemed to work. Berlin was saved, and Greece and Turkey protected. In 1950, however, after South Korea was invaded, the Chinese were not deterred from intervening on behalf of North Korea. By the time Eisenhower took office in 1952, American credibility was ebbing. Making matters worse, Eisenhower – who had campaigned on a pledge to end the war – publicly vowed to avoid future land wars in Asia. Evidently, the US threat to resist communist expansion anywhere and any time could no longer be considered as given.

The Eisenhower administration's response was its New Look defense

4. We distinguish between actual and perceived credibility. When information about preferences is complete, actual and perceived credibility are the same. Under incomplete information, they will differ. In the latter case, it is the perception of this preference relationship that becomes critical.

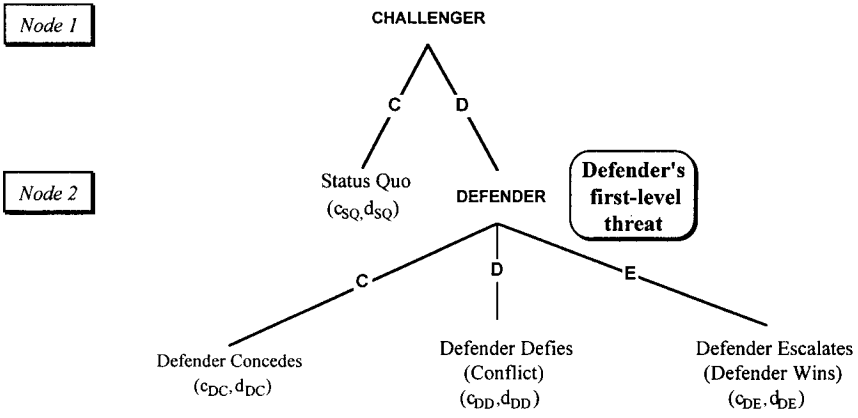
5. Under *incomplete information* deterrence works as long as Defender's threat is credible enough.

policy that de-emphasized conventional forces and relied instead on atomic and, later, nuclear weapons to protect the status quo. At the heart of the New Look was the doctrine of Massive Retaliation. As enunciated by Secretary of State John Foster Dulles, the US notion of Massive Retaliation depended 'primarily on a great capacity to retaliate, instantly, by means and at places of our own choosing'.

The idea was to deter the Soviet Union by threatening to transform local conflicts into strategic confrontations. Since the US maintained a distinct strategic advantage over the Soviet Union in 1952, this threat was inherently more credible than a threat to intervene in a more limited way in a peripheral area.

In game-theoretic terms, the New Look sought to transform the game of Figure 1 to that of Figure 2, wherein Defender has a third response option, to *escalate* (E), giving rise to an additional outcome: *Defender Escalates* (DE). Presumably, unilateral escalation would lead to a victory for Defender (i.e. the United States) while a non-escalatory response-in-kind (i.e. a choice of D) would result in a crisis or some other kind of limited conflict. Note that even with the addition of Defender's escalatory option, a conflict spiral remains impossible since Challenger has no opportunity to counter-escalate.

Underlying the need for Dulles' new strategic doctrine, and its wisdom,



Key: C = Cooperate/Concede; D = Demand/Defy; E = Escalate

Additional assumptions: Challenger: DD > DE (Reflects costs of escalated conflict)
Defender: DE > DC and DD (Dulles' interpretation of MR)

Figure 2. Massive Retaliation circa 1954

are a number of assumed preference relationships. First, the threat of Massive Retaliation is unnecessary unless Defender's threat to respond-in-kind is seen to lack credibility. Hence, when analyzing this particular policy, we assume Defender prefers DC to DD (Zagare and Kilgour, 1993b).[6] Similarly, Massive Retaliation as a strategic doctrine is incoherent unless Defender prefers to escalate unilaterally, i.e. prefers DE to DC. Thus, for now, we presume Defender prefers DE to DC, and DC to DD. As well, we presume Challenger prefers *Limited Conflict* (DD) to Defender Escalates (DE). After all, limited conflicts take place on Challenger's terms, while any conflict that occurs after Defender escalates unilaterally would be on Defender's terms.

Given these assumptions, it is easy to understand the reasoning behind Massive Retaliation: the status quo is stable, and deterrence works. The logic was impeccable. Unfortunately, the underlying game form proved ephemeral. Before long, critics like William Kaufmann (1956: 21) were charging that 'if we are challenged to fulfill the threat of massive retaliation, we will be likely to suffer costs as great as those we inflict'. In other words, the underlying game was about to change: the Soviet Union was thought to be capable of fully responding to any strategic attack by the US.

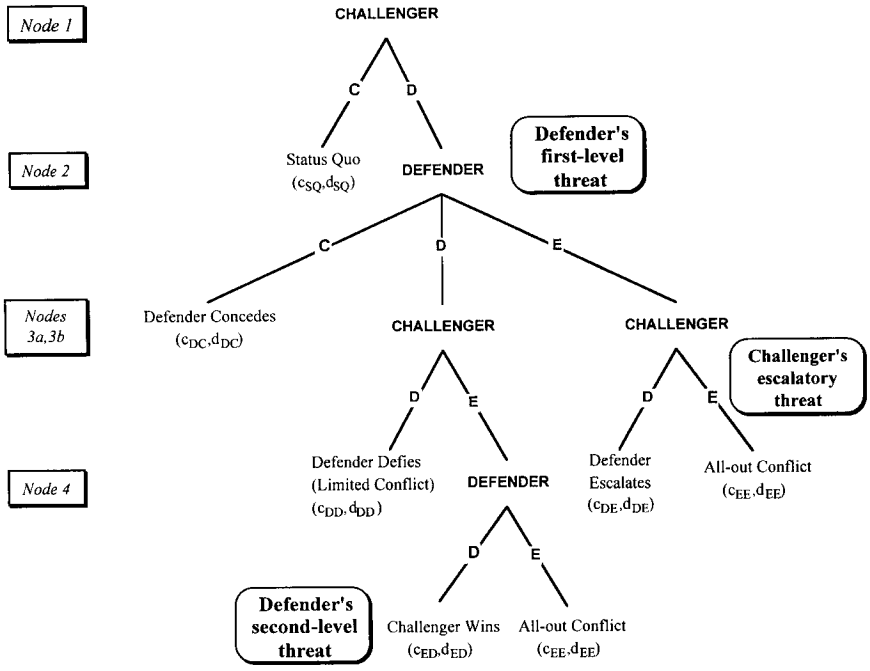
This important development is reflected in the game tree of Figure 3, which we call the Asymmetric Escalation Game.[7] Note Challenger's option to escalate or not at Node 3a, and its option to (counter-)escalate at Node 3b; note as well Defender's option to counter-escalate at Node 4. These additional choices give rise to three more outcomes. Specifically, if one player escalates and the other does not, the player that escalates gains an advantage [either Defender Escalates (DE) or Challenger Wins (ED)]. If both escalate, All-out Conflict (outcome EE) occurs.[8]

The expanded set of choices also introduces two additional threats into the Asymmetric Escalation Game. Challenger now has a threat to counter-escalate at Node 3b. Defender, in addition to its first-level threat to respond-in-kind (i.e. choose D) at Node 2 should Challenger demand a

6. In terms of the game of Figure 1, we presume Defender is Soft. In the subsequent analysis we relax this assumption. We make it here only for the limited purpose of discussing the historical underpinnings of our model.

7. While we develop this model in the context of the superpower rivalry, we believe that its underlying dynamic is representative of a large number of other non-nuclear extended deterrence relationships. Our conclusions, therefore, apply to a wide variety of real world interactions.

8. Admittedly, our two-level game does not provide for the possibility of a very long conflict spiral, but it is a spiral nonetheless. Based on a comparative evaluation of two-level (Zagare, 1990b) and three-level escalation games (Zagare, 1992) of complete information, however, we offer the conjecture that the general conclusions we draw are not particularly sensitive to the precise number of moves in the game.



Key: C = Cooperate/Concede; D = Demand/Defy; E = Escalate

Figure 3. Asymmetric Escalation Game
Source: Zagare and Kilgour, 1995, 1996

change in the status quo, now has a *second-level threat*: to counter-escalate at Node 4 should Challenger escalate first at Node 3a.[9] As one might expect, these additional threats play an important role in stabilizing or destabilizing the status quo in the Asymmetric Escalation Game; they are also key determinants of *intra-war* deterrence (Schelling, 1966). We assume that both threats are capable in the sense discussed above (i.e. each player prefers Limited Conflict (DD) to All-out Conflict (EE)).

For the first time in our model, conflict spirals are a distinct possibility. Both players can make choices that culminate in disaster. It is no accident that it was around this time (i.e. the mid-1950s) that the strategic literature on conflict escalation began to bifurcate (Smoke, 1977: Ch. 2). Classical deterrence theorists, fixating on stability, modified their analyses to take

9. Depending upon the specific empirical referent we have in mind at a particular moment, we may also refer to these two choices either as Defender's tactical (or sub-strategic) and strategic level threats or as Defender's conventional and nuclear threats.

account of the evolving realities. For their part, however, spiral theorists developed what were seen as competing models that highlighted action–reaction processes frequently culminating in disaster (e.g. Richardson, 1960; Holsti et al., 1968; Pruitt, 1969).

We concede that in devising the Asymmetric Escalation Game we have, somewhat arbitrarily, restricted the choices available to the players once both choose D and Limited Conflict ensues, and when both escalate and an All-out Conflict occurs. This does not mean that other choices are impossible, either theoretically or empirically, but merely that we have (conceptually) folded these choices into the choices leading to outcomes DD and EE. For instance, a limited conflict that persists in an equilibrium state could, eventually, evolve into a prolonged stalemate or a chronic crisis if both players hold firm, a clear victory for one of the players after the other backs down, a negotiated settlement if the players decide to mediate their differences, and so on. And All-out Conflicts eventually end when one or both states decide that it is no longer in their interest to continue. But we ignore these complexities to focus exclusively on conflict spirals and the dynamics of deterrence. To do otherwise would unduly complicate our model with no sure prospect of a commensurate analytical payoff.

We do not believe that these simplifications affect our conclusions in any way. Our focus is on explaining how wars and other all-out conflicts begin, not how they end, when deterrence fails and when conflict spirals occur. Once a war erupts, choices that may subsequently become available are largely beside our point. And even if removed, the conditions that precipitate a conflict may be irrelevant to ending it.

For example, many historians (e.g. Kagan, 1995) believe that the absence of a credible British threat contributed mightily to the German decision to invade Belgium in 1914. Obviously the Germans knew early on that the British preferred to fight rather than back off, but once they found this out the point was moot. The war had begun because of a belief to the contrary, and the participants then had to find a way out of their predicament. To be sure, either side could have backed off once the hostilities escalated, but neither did. Both clearly preferred fighting on to capitulating. But even if the Germans did not, adding this additional choice to our game tree would not change our analysis of either why deterrence had failed or why the conflict had initially spiraled out of control.

In the exploration of the Asymmetric Escalation Game, we assume that the players prefer winning to losing. To reflect the costs of conflict, we also assume that the players prefer to win or, if it comes to it, lose, at the lowest level of conflict. Thus Challenger prefers Defender Concedes (DC) to Challenger Wins (ED) – and so does Defender.

As before, we leave open each player's preference for executing its threat(s). This means that Challenger may be one of two types (Hard and

Soft) and that Defender may be one of four: Hard at the first level but Soft at the second (i.e. type H/S); of type S/H: Soft at the first level but Hard at the second; of type H/H: Hard at both levels; or of type S/S: Soft at both. In the subsequent analysis we assume that each player knows its own type and has probabilistic knowledge of the type of its opponent. A player's probability of being Hard is taken as a measure of its credibility.

In sum, we make the following assumptions about the players' preferences:

Challenger
Defender Concedes > *Status Quo* > *Challenger Wins* > *Limited Conflict* > [*Defender Wins* and *All-out Conflict*]

Defender
Status Quo > *Defender Wins* > [*Defender Concedes* and *Limited Conflict*] > [*Challenger Wins* and *All-out Conflict*],

where '>' means 'is preferred to'. The relative preferences for those outcomes enclosed in brackets are left open, i.e. are the parameters of our model.

2. Incomplete Information

To delineate the conditions associated with successful deterrence, limited conflicts and escalation spirals, we examine the Asymmetric Escalation Game under incomplete information. We assume incomplete information because, without it, war is unlikely (Blainey, 1988). In our model, the principal source of uncertainty is each player's lack of information about the other's preference for conflict over capitulation at a particular level of play, i.e. the credibility of the other player's threat(s).[10]

Our definition allows for several conceptually compatible ways to interpret uncertainty. For example, not knowing an opponent's preference between conflict and capitulation implies uncertainty about an opponent's intention to execute a threat. This is conceptually equivalent to saying that there is uncertainty about likely choices at several nodes of the game tree or, in the context of our model, about 'intra-war' behavior. Similarly, lack of knowledge about an opponent's preferences between conflict and capitulation could stem from not knowing the value an opponent places on the issue in dispute (is it worth fighting for?) or the costs it attaches to conflict.

In real world deterrence games, of course, there may be other sources of

10. For an extended discussion of the connection between credibility and uncertainty, see Kilgour and Zagare (1991).

uncertainty. For instance, one or both players might have private information about its capability to wage war (Morrow, 1989a, b). Or the players might be uncertain about the consequences of their choices (Schelling, 1960; Nalebuff, 1986; Powell, 1990).[11] However, we ignore these other potential sources of uncertainty, not because they are unimportant, but to focus solely on the critical role played by credible threats in deterring or capping conflict or contributing to an escalatory spiral. Credible threats in the sense defined above have not been extensively analyzed in the formal strategic literature. Like Smoke (1977: 4) we believe that 'the decision to escalate ... is a strategic issue, involving not only assessment of the immediate advantage to one's own side, but also difficult and often painfully uncertain calculation of the possibilities for counterescalation by the enemy'.

To model incomplete information about an opponent's type, we assume the utilities to Defender at outcome DD (D_{DD}) and to both players at outcome EE (D_{EE} and C_{EE}) are binary random variables (indicated by upper case letters) with known distributions. Specifically, we assume the following to be common knowledge:

$$C_{EE} = \begin{cases} c_{EE+} \text{ with probability } p_{Ch} \\ c_{EE-} \text{ with probability } 1 - p_{Ch} \end{cases}$$

$$(D_{DD}, D_{EE}) = \begin{cases} (d_{DD+}, d_{EE+}) \text{ with probability } p_{HH} \\ (d_{DD+}, d_{EE-}) \text{ with probability } p_{HS} \\ (d_{DD-}, d_{EE+}) \text{ with probability } p_{SH} \\ (d_{DD-}, d_{EE-}) \text{ with probability } p_{SS} \end{cases}$$

where $d_{DD+} > d_{DC} > d_{DD-}$, $d_{EE+} > d_{ED} > d_{EE-}$, and $c_{EE+} > c_{DE} > c_{EE-}$; $0 < p_{HH} < 1$, $0 < p_{HS} < 1$, $0 < p_{SH} < 1$, $0 < p_{SS} < 1$; $p_{HH} + p_{HS} + p_{SH} + p_{SS} = 1$, and $0 < p_{Ch} < 1$.

In words, we assume that Defender believes Challenger to be *Hard* with probability p_{Ch} and *Soft* with probability $1 - p_{Ch}$. Likewise, Challenger believes Defender to be of type HH with probability p_{HH}, of type HS with probability p_{HS}, of type SH with probability p_{SH}, and of type SS with probability p_{SS}. These beliefs are common knowledge.

These beliefs are also taken as a measure of threat credibility. For example, the greater the value of p_{Ch} (i.e. Defender's belief that Challenger is *Hard*), the greater the perceived credibility of Challenger's threat. Similarly, the greater the value of p_{HH}, the greater the perceived credibility of both of Defender's threats – first and second level.[12]

The overall probability that Defender prefers conflict to capitulation at

11. For discussions of this literature, see Zagare (1990), O'Neill (1994), and Carlson (1995).

12. See footnote 4.

Table 1. Type and Credibility Parameters

Variable	Probability that	Measures the Perceived Credibility of
p_{Ch}	Challenger is Hard	Challenger's threat (Node 3b)
$1 - p_{Ch}$	Challenger is Soft	
p_{HH}	Defender is of type HH	Defender's first- *and* second-level threats
p_{HS}	Defender is of type HS	
p_{SH}	Defender is of type SH	
p_{SS}	Defender is of type SS	
p_{Tac}	Defender is of type HH or HS	Defender's first-level threat (Node 2)
$1 - p_{Tac}$	Defender is of type SH or SS	
p_{Str}	Defender is of type HH or SH	Defender's second-level threat (Node 4)
$1 - p_{Str}$	Defender is of type HS or SS	

the first (or tactical) level is the perceived credibility of Defender's first-level threat. This probability, that Defender is of type HH *or* type HS, is denoted $p_{Tac} = p_{HH} + p_{HS}$; therefore, the overall probability that Defender prefers capitulation to conflict at the first level is $1 - p_{Tac} = p_{SH} + p_{SS}$. Similarly, the perceived credibility of Defender's second-level (or strategic) threat is $p_{Str} = p_{HH} + p_{SH}$; so $1 - p_{Str} = p_{HS} + p_{SS}$ is the probability that Defender prefers capitulation to conflict at the second level. Table 1 summarizes our notation for players' types and their perceived credibilities.[13]

3. Behavioral Possibilities

In a game of incomplete information, all rational behavioral possibilities are captured by the set of Perfect Bayesian Equilibria. A Perfect Bayesian Equilibrium is a set of strategies and beliefs, one for each player in the game. Given its beliefs, no one player can expect to do better by switching, unilaterally, to a different strategy. Furthermore, each player rationally updates its beliefs (about its opponent's type) whenever it observes the opponent's action choices (Fudenberg and Tirole, 1991). Thus, the outcomes supported by Perfect Bayesian Equilibria are precisely the outcomes that can be expected in rational play. Our view is that only Perfect Bayesian Equilibria will evolve and persist in the Asymmetric Escalation Game with incomplete information.

Knowing the Perfect Bayesian Equilibria and the conditions under which

13. Henceforth we may drop the adjective 'perceived' when dealing with perceptions of preference relationships.

they exist, then, is quite informative. This information reveals if and when deterrence will succeed and, if it fails, how. Conversely, the conditions associated with escalation spirals (or other behavioral sequences) can also be specified. In other words, the Perfect Bayesian Equilibria are a theoretical tool for addressing Jervis's (1976: 96) question concerning 'the conditions under which one model rather than the other is appropriate'. Parenthetically, we note that by distinguishing outcomes according to the *belief systems* that give rise to them, we provide an answer to Jervis' question in precisely the terms it was posed.

Before proceeding, however, a few caveats are in order. First, for the purpose of this essay, we do *not* define deterrence merely as the possible survival of the status quo or of some other outcome. The status quo does occasionally persist under most Perfect Bayesian Equilibria of the Asymmetric Escalation Game with incomplete information. (For the details, see Zagare and Kilgour, 1996.) In many of these cases, however, the possibility is remote. Thus we identify successful deterrence only with those equilibria under which one player is generally dissuaded from taking an action leading to an immediately better outcome precisely because it fears the other will retaliate – either in-kind or by escalating.

Despite this restriction, we are able to distinguish several distinct patterns of deterrence. The first, the traditional notion, is that deterrence is established when Challenger has *no* motivation to demand an alteration of the status quo. By contrast, the second and third patterns require the non-standard argument that deterrence remains relevant even after a conflict has erupted (Snyder, 1961; Schelling, 1966: 191). In a limited war, for example, each side might choose not to escalate precisely because it fears the other will counter-escalate. Or it could be the case that one player (Defender) decides not to respond after the other (Challenger) initiates. Clearly, deterrence is operative in this instance as well. Thus we consider both *pre-war* and *intra-war* behavioral sequences to be potentially consistent with the precepts of classical deterrence theory.

Second, we do not consider conflict spirals and escalation to be equivalent. While some conflicts escalate immediately to the highest level from the very onset of hostilities, others reach an acute stage after a long series of moves and countermoves. The 1973 war in the Middle East, which began with a concerted surprise attack by Egypt and Syria against Israel, illustrates the former. The First World War is the prototype of the latter; what began as a minor incident in the Balkans slowly, deliberately, and perhaps inexorably, spiraled out of control as ultimata were followed by mobilization plans, alerts, counter-alerts, frontal attacks and, eventually, counter-attacks. More than simply the escalation of conflicts to war, we hope to explain and place in context this classic escalation spiral.

Finally, we restrict our analysis to the special case in which Challenger is

probably Hard, i.e. its threat to counter-escalate is highly credible. There are two reasons for this focus, one technical and the other theoretical. Technically, when Challenger is likely Hard, the equilibrium structure of the Asymmetric Escalation Game with incomplete information is simple and less subject to minor but complex exceptions than when Challenger is likely Soft. A more important reason, however, is that this is the more interesting case. *Ceteris paribus*, deterrence is more likely, and conflict spirals less likely, when Challenger is probably Soft. In other words, the real test for proponents of deterrence occurs when Challenger is seen to be likely willing to run the risk of war. As well, this is precisely the circumstance that spiral theorists argue is most prone to deterrence failures and conflict spirals. Thus, by focussing attention on the most problematic case, we accentuate the theoretical distinctions between deterrence and spiral theorists.

4. Deterrence and Conflict Spirals

When does traditional deterrence succeed? Under what conditions can the escalation process be contained? When will conflict spirals occur? In this section we attempt to answer these questions by considering the strategic characteristics of the three groups of Perfect Bayesian Equilibria that can exist when Challenger is likely Hard:

1. *Deterrence Equilibria* that depend on the threat of escalation;
2. the *No-Response Equilibrium*; and
3. the *Spiral Group* of four equilibria that includes two additional forms of *Deterrence Equilibria*, a *Constrained Limited-Response Equilibrium* and an *Escalatory Limited-Response Equilibrium*.

Deterrence Equilibria are associated with traditional notions of deterrence success; the No-Response and the Constrained Limited-Response Equilibria with *intra*-war deterrence; and the Escalatory Limited-Response Equilibria with conflict spirals and reciprocated levels of violence. We begin by describing the strategic properties of the various equilibria (see Table 2). Subsequently we address their implications for deterrence and spiral theories.

Traditional Deterrence

Traditional deterrence can evolve in three very different ways in the Asymmetric Escalation Game with incomplete information. Regardless of the path to deterrence, however, Challenger's action choice is always the same: regardless of its type, Challenger never initiates and the outcome of

Table 2. Equilibria of the Asymmetric Escalation Game when Challenger has High Credibility

	Challenger					Defender						
	x		w		q_{HH}	y		z				r
	x_H	x_S	w_H	w_S		y_{HH}	y_{HS}	z_{HH}	z_{HS}	z_{SH}	z_{SS}	
Deterrence (typical)												
Det_1	0	0	1	1	*Small*	0	0	1	1	1	1	$\leq d_1$
No-Response												
NRE	1	1	1	1	$\leq c_q$	0	0	0	0	0	0	p_{Ch}
Spiral Family												
Det_2	0	0	0	0	$p_{Str/Tac}$	1	1	0	0	0	0	$\geq d_2$
Det_3	0	0	d^*/r	0	c_q	1	v	0	0	0	0	$\geq d_2$
$CLRE_1$	1	1	0	0	$p_{Str/Tac}$	1	1	0	0	0	0	p_{Ch}
$ELRE_3$	1	1	d^*/p_{Ch}	0	c_q	1	v	0	0	0	0	p_{Ch}

Note: Table 2 is excerpted from Table 1 of Zagare and Kilgour (1996) which should be consulted for details of definitions and interpretations. A brief summary of the strategic and belief variables appearing in Table 2 is given here for convenience.

The probability that Challenger initiates (i.e. chooses D) at Node 1 of the game of Figure 3 is denoted x. In fact, this probability can depend on Challenger's type – if Challenger is Hard, the initiation probability is x_H; if Soft x_S. Likewise w_H and w_S are the probabilities that Hard and Soft Challengers, respectively, escalate (i.e. choose E) at Node 3a. At Node 3b, Challenger always chooses E if Hard and D if Soft.

Similarly, Defender chooses D at Node 2 with probability y, E with probability z, and C with probability $1 - y - z$. Again, these probabilities can depend on Defender's type, so they are denoted y_{HH}, z_{HS}, etc. It can be proven that $y_{SH} = y_{SS} = 0$. At Node 4, Defender chooses E if strategically Hard (type HH or SH), and chooses D otherwise.

Finally, players revise their initial probabilities about their opponent's type as they observe the opponent's actions. Of these revised probabilities, the only two that are important to the equilibria are those shown in Table 2. Defender's revised probability that Challenger is Hard, given that Challenger initiates, is denoted r. Defender's revised probability that Challenger is of type HH, given that Defender chooses D (response-in-kind) at Node 2, is denoted q_{HH}.

the game is always the Status Quo. What distinguishes the various Deterrence Equilibria are Challenger's and Defender's intentions 'off the equilibrium path'. These intentions reflect the players' beliefs about each other's type and their planned choices at nodes (or decision points) that are not reached because deterrence is successful.

The first group of deterrence equilibria is a family of several Perfect Bayesian Equilibria that we call the *Challenger-Soft Deterrence Equilibria.*

Typical of this family is Det_1 as shown in Table 2. Under Challenger-Soft Deterrence Equilibria at least some types of Defender intend to escalate by choosing E at Node 2. Thus, the traditional deterrence that emerges from a Challenger-Soft Deterrence Equilibrium depends on Defender's willingness to escalate first. (Under Det_1, the most extreme Challenger-Soft Deterrence Equilibrium, all types of Defenders escalate with certainty at Node 2.)

For a member of this family to exist, Defender must believe that any demand for a change in the status quo would be a mistake made by a genuinely Soft Challenger.[14] In other words, for a Challenger-Soft Deterrence Equilibrium to come into play, Defender must believe that Challenger is unlikely to be Hard *even should Challenger initiate a conflict.*[15] Only this particular belief would rationally support Defender's intention to escalate at Node 2. In turn, Defender's intention to escalate first deters Challenger from upsetting the status quo.

Provided Defender initially believes Challenger to be likely Soft, this subtle interplay of beliefs and action choices seems plausible enough. On the other hand, it is difficult to imagine a situation in which Defender, after observing an act of unprovoked hostility, concludes that a Challenger – who was originally thought to be probably Hard – is actually Soft. For this reason we consider the behavioral pattern associated with *Challenger-Soft Deterrence Equilibria* to be unlikely under the special circumstances explored in this essay (i.e. when Challenger is likely Hard) and, hence, irrelevant to the debate between Deterrence and Spiral Theorists. Other equilibria of the Challenger-Soft group share this characteristic – to support rationally the intention to escalate, Defender must interpret any unanticipated initiation by Challenger as a sign that Challenger is likely Soft. Thus all equilibria of this group are implausible in the same way.[16]

By contrast, the *Defender-Hard Deterrence Equilibrium* (or Det_2) is a plausible outcome of the Asymmetric Escalation Game under study. Unlike the Challenger-Soft Deterrence Equilibrium, the Defender-Hard Deterrence Equilibrium does not require Defender to escalate first. In fact,

14. Det_1 is actually independent of Challenger's a priori beliefs about Defender's credibility; other members of the family, by contrast, place certain restrictions on Challenger's beliefs. For more details, see Zagare and Kilgour (1996).

15. More technically, Defender's updated belief (probability r) that Challenger is Hard given that Challenger chooses D at Node 1 must be relatively small.

16. This is not to say that the Challenger-Soft Deterrence Equilibrium is implausible under all circumstances. In fact, the Challenger-Soft Deterrence Equilibrium may help explain both the stability of the superpower relationship during the Eisenhower administration, and why, despite crises in the Balkans in 1905, in 1908, and again in 1912, war did not break out in Europe until 1914. After the turn of the century, Great Britain – like the US during the 1950s – relied mainly on an escalatory threat to deter its principal rival (Massie, 1991; Kagan, 1995).

the form of traditional deterrence that emerges under Det_2 rests *entirely* on the more limited threat of responding-in-kind at Node 2.

The existence of a Defender-Hard Deterrence Equilibrium depends solely on Challenger's beliefs about Defender's type. (Defender's a priori beliefs are completely immaterial to the existence of Det_2.) Specifically, for Det_2 to exist, both Defender's first- and second-level threats must be highly credible: Challenger must believe it quite likely that Defender is tactically Hard, and given that Defender is tactically Hard, Challenger must place a fairly high probability on Defender being strategically Hard also.

Given these beliefs, Challenger intends not to escalate at Node 3a because it believes that Defender will likely counter-escalate at Node 4; and because Challenger believes that Defender will almost certainly respond-in-kind at Node 2 – thereby forcing Challenger to back down at Node 3a – Challenger decides *not* to initiate at Node 1.

Although the final Deterrence Equilibrium, Det_3, is also a plausible outcome of the Asymmetric Escalation Game, it is not likely. As explained in the next section, the conditions under which Det_3 exists are quite restricted. Nonetheless, because it is closely linked to Det_2, and an integral component of the Spiral Family, Det_3 remains a theoretical possibility worth describing.

Det_2 and Det_3 are the only deterrence equilibria that depend completely on Defender's threat to respond-in-kind to deter Challenger. As well, each requires Defender to be likely Hard at the second level, given it is Hard at the first level of play. Under Det_3, however, this conditional probability is somewhat lower.

In terms of action choices at Det_3, Defender plans to respond-in-kind with certainty if it is of type HH, and probabilistically if of type HS; otherwise Defender does not respond at all (i.e. it capitulates at Node 2). Since the conditional probability that Defender is Hard at the second level, given that it is Hard at the first, is slightly lower under Det_3 than under Det_2, a Hard Challenger will intend to escalate probabilistically at Node 3a. It is the willingness of a Defender who is HS to respond-in-kind sometimes that permits a Hard Challenger to risk escalating sometimes, and contrariwise. In the end, Challenger is deterred. Keep in mind that for this delicate balancing act to take place, both of Defender's threats must be fairly credible, that is, Defender must be likely to be tactically Hard, and moderately likely to be strategically Hard as well.

Non-traditional Deterrence

As already noted, deterrence can still operate even after the status quo has been violated. Crises that do not erupt into open hostilities, cold wars that do not turn hot, unilateral acts of aggression, and limited conflicts that do

not escalate illustrate that traditional deterrence can break down in a way that respects some limits. True – deterrence fails. But on another level, the absence of all-out conflict signals both a modicum of restraint and circumscribed success: in each situation at least one player eschews an action leading to an immediately better outcome precisely because it fears the other's response.

Non-traditional deterrence can occur in two distinct ways in the variant of the Asymmetric Escalation Game explored in this essay. In each case, the status quo is upset: Challenger, whether Hard or Soft, simply initiates at Node 1. Defender's action choices, however, depend on its type and on which of two Perfect Bayesian Equilibria is in play.

Under the *No-Response Equilibrium*, Defender simply capitulates – as the French did after Hitler occupied and remilitarized the Rhineland in 1936. Defender gives in (i.e. is deterred from responding-in-kind and from escalating) because Challenger is very likely Hard and, therefore, prone to escalate at Node 3a or to counter-escalate at Node 3b. To support its choice at Node 3a, however, Challenger must believe that a Defender who unexpectedly *responds-in-kind* at Node 2 is more likely to be of type HS than of type HH. We find this to be a plausible belief since, *ceteris paribus*, type HH Defenders are more likely to *escalate* than type HS Defenders.

By contrast, Defender's action choices under $CLRE_1$ – the only form of *Constrained Limited-Response Equilibrium* that exists when Challenger is likely Hard – involve a response-in-kind for certain, but only when Defender is of type HH or HS.[17] Otherwise, Defender capitulates. In fact, *Defender Concedes* is the most likely outcome of play under $CLRE_1$ since this member of the Spiral Family of Perfect Bayesian Equilibria exists when Defender is likely Soft at the first level, i.e. when p_{Tac} is low and p_{HS} is not too large. Thus, when Challenger chooses D at Node 1, it does so with the expectation that its demands will almost certainly be met.

Put in another way, a response-in-kind will surprise Challenger under $CLRE_1$. In this unlikely event, Challenger will be forced to update its beliefs about Defender's type. Clearly, Challenger will conclude that Defender is of type HH or HS, since only Defenders of these two types can rationally choose D at Node 2. Moreover, under any Constrained Limited-Response Equilibrium including $CLRE_1$, *if* Defender is Hard at the first level, it is also likely to be Hard at the second level as well, i.e. more likely to be of type HH than of type HS. Fearing this possibility, Challenger is, understandably, completely deterred from escalating at Node 3a; instead, it always chooses D at Node 3a, settling for limited conflict.

17. There are five distinct forms of Constrained Limited-Response Equilibria. The remaining four forms exist only when Challenger's credibility falls below a certain threshold. For details, see Zagare and Kilgour (1996).

In sum, while the status quo does not survive under either a No-Response Equilibrium or under $CLRE_1$, deterrence still plays an important role in the Asymmetric Escalation Game with incomplete information when either of these non-traditional deterrence equilibria is in play. In the first instance, deterrence is asymmetric: precisely because *Defender* is deterred, Challenger is able to initiate with impunity. In the second instance deterrence is more symmetric: each player is able to deter *completely*, but only to deter from escalating. As a consequence, limited conflicts may evolve under $CLRE_1$, and if so, they *always* remain limited.

Conflict Spirals

Such is not the case, however, under $ELRE_3$, the only form of *Escalatory Limited-Response Equilibrium* that exists when Challenger is likely Hard. While it is *possible* for a limited conflict (outcome DD) to occur under $ELRE_3$, such a dénouement is, at best, a remote possibility. In fact, the most likely outcome of a game played under this spiral equilibrium is, once again, DC, Defender Concedes.

As with the No-Response Equilibrium and $CLRE_1$, Challenger, whatever its type, always chooses D at Node 1, thereby upsetting the status quo. What happens next depends on Defender's type. Under $ELRE_3$, Defender is likely to be of either type SS or SH. Such Defenders *always* concede at Node 2.[18] In the less likely event that Defender is Hard at the first level, Defender will respond-in-kind with certainty if it is also Hard at the second level (i.e. of type HH) and probabilistically if it is Soft at the second level (i.e. of type HS). Given the probabilities, however, a response-in-kind will once again surprise Challenger.

Up to this point of surprise, behavior and expectations are similar under $ELRE_3$ and $CLRE_1$. What separates these two equilibria are Challenger's expectations should Defender unexpectedly choose D at Node 2. Recall that under $CLRE_1$, Defender will only respond-in-kind if Hard at the first level; and if Defender is Hard at the first level, then it is likely Hard at the second level as well. This is why Challengers never escalate first under a Constrained Limited-Response Equilibrium.

Under $ELRE_3$, though, a Defender that responds-in-kind is much more likely to be of type HS than of type HH. For this reason, a Hard Challenger, the object of our attention, simply escalates at Node 3a. If it so happens that Defender is actually Hard, the heretofore limited conflict spirals out of control.

18. This is why Defender Concedes is the most likely outcome under any Escalatory Limited-Response Equilibrium.

5. Discussion

We now return to our original questions: When does deterrence work? Under what conditions are limited conflicts possible? When will conflicts seem to take on a life on their own by escalating out of control? Obviously, our model is too simple to be definitive, but it does suggest answers.

Given our fixed assumption that Challenger is likely to prefer to escalate at Node 3b, it should not be surprising that our answers are in terms of the main parameters of the model: Defender's perceived credibilities. In fact, another way to pose these questions is: what kind of commitment must Defender be seen to have to deter conflict altogether, or to prevent low-level conflicts from escalating, given it is likely that Challenger considers the stakes worth fighting for?

To answer this question, we next consider the existence conditions associated with each possible Perfect Bayesian Equilibrium of the Asymmetric Escalation Game with incomplete information. As already noted, the Perfect Bayesian Equilibria constitute all predictions of the model that are consistent with rational behavior.

We begin by noting that Perfect Bayesian Equilibria of the Challenger-Soft family and the No-Response Equilibrium *always* exist no matter what Defender's credibilities. But since Det_1 and other members of the Challenger-Soft Deterrence Equilibrium family are based on beliefs that are implausible, particularly so given our presumption about Challenger's credibilities, we do not consider them viable solutions to the Asymmetric Escalation Game. The No-Response Equilibrium, however, is not so easily dismissed. As long as Challenger is likely Hard, it will *always* exist as a logical possibility along with *precisely one* of the Spiral Family of Perfect Bayesian Equilibria: Det_2, Det_3, CLRE_1 and ELRE_3. Which of these four equilibria will exist is determined by the perceived credibilities of Defender's first- and second-level threats.

Exactly how Defender's credibilities determine which equilibrium of this 'gang of four' exists is depicted in three-dimensional space in Figure 4. Every possible combination of Defender's credibilities is represented as a point in the tetrahedron shown in the center of this figure. The right horizontal axis represents the probability that Defender is of type HH, the lower-left (horizontal) axis the probability that Defender is of type SH, and the vertical axis the probability that Defender is of type HS. Thus, any point in the three-dimensional space represents a combination of non-negative values of p_{HH}, p_{HS} and p_{SH} with a sum less than or equal to 1. In fact, p_{SS} equals the difference between this sum and 1, which is also the distance between the point (p_{HH}, p_{HS}, p_{SH}) and the front face of the tetrahedron. For example, the point (0,0,0) represents the combination $p_{\text{HH}} = p_{\text{HS}} = p_{\text{SH}} = 0, p_{\text{SS}} = 1$.

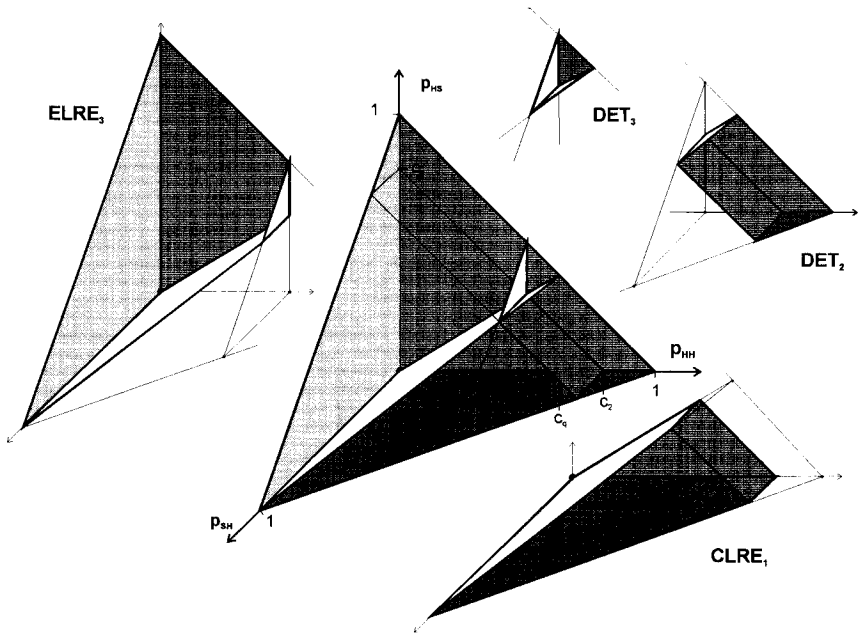


Figure 4. Existence Regions for Equilibria of the Spiral Family

On a more intuitive level, Figure 4 can be visualized as a corner of a room with two walls and a floor, all at right angles, with the fourth face being a downward sloping plane. The side wall is light gray, the back wall is medium gray, while the floor is dark gray. Of course, to enable us to peer into this corner room, the front face must remain transparent.

As Figure 4 suggests, for traditional deterrence to have a chance, *both* of Defender's threats must be fairly credible. Thus, the two closely related Deterrence Equilibria, Det_2 and Det_3, occupy a small region in the right-hand side of the tetrahedron, where p_{HH} is large, p_{HS} is not too large, and p_{SH} and p_{SS} are small. Defender is likely tactically Hard; this explains its propensity under either Det_2 or Det_3 to respond-in-kind at node 2, whatever its actual type. But this tendency alone is not sufficient to deter Challenger. Defender's willingness to respond-in-kind also rests on its ability to dissuade Challenger from escalating at Node 3a. For this to occur, Defender's second-level threat must be highly credible as well; in other words, for deterrence to evolve under either Det_2 or Det_3, Defender must likely be both strategically and tactically Hard – i.e. p_{HH} must be large.

A somewhat different behavioral pattern emerges, however, when the credibility of Defender's first-level threat is just slightly too small to sustain either Deterrence Equilibrium. This is the region of $CLRE_1$ which exists as

a forward-leaning wedge running from the left side wall to the front face of the tetrahedron (lower right of Figure 4).

The small reduction in the credibility of Defender's first-level threat provides even a Soft Challenger with an incentive to initiate at Node 1. After all, Defender believes that Challenger is likely Hard and is therefore completely deterred from escalating first, even if it is of type HH. Thus under $CLRE_1$ Challenger banks on Defender's preference for Defender Concedes over Limited Conflict and takes decisive action. Often Challenger's gamble pays off and Defender capitulates. From time to time, however, Challenger guesses wrong and Defender reacts.

Defender's response-in-kind is Challenger's first clue that Defender is Harder than it might have hoped since only a Defender who is tactically Hard, i.e. prefers Limited Conflict over Defender Concedes, would rationally choose D at Node 2. But it is the conclusion that Challenger draws from Defender's unexpected response that is the distinguishing feature of $CLRE_1$.

Notice from Figure 4 that the upper face of the $CLRE_1$ region slopes upward away from the bottom edge of the left side wall. At $CLRE_1$, the probability that Defender is of type HH is never very large (maximum c_2). However this sloping 'ceiling' means that the probability that Defender is of type HS is always small *relative to the probability that it is of type HH*. In consequence, given that Defender has already demonstrated that it is tactically Hard (HH or HS) by responding-in-kind, there is a relatively high probability that it is also strategically Hard (HH rather than HS) and would counter-escalate at Node 4. This probability is high enough to convince Challenger, whatever its type, never to escalate at Node 3a.

Conflict spirals occur precisely when these conditions are not satisfied. Notice from Figure 4 that under $ELRE_3$ Defender is less likely to be of type HH than under $CLRE_1$, and much more likely to be of type HS than of type HH. Confident that Defender will never escalate, both types of Challenger initiate, again with the expectation that all demands will probably be met. Normally, Challenger will not be disappointed.

As with play under $CLRE_1$, however, over time Challenger is likely to face measured resistance from a Defender with a preference for Limited Conflict over Defender Concedes (i.e. a Defender of type HS or HH). In the unlikely event that Challenger is Soft, it will choose not to escalate and a limited conflict will ensue. But in the much more likely event that Challenger is Hard, it *may* escalate precisely because Defender is unlikely to counter-escalate at Node 4. As Figure 4 shows, under $ELRE_3$, a tactically Hard Defender is *less* likely to be of type HH than of type HS.

At this point, Defender will back off – but only if Challenger has guessed correctly. A Defender of type HH simply counter-escalates at Node 4. If so, the result will be tragic, the spiral complete. This, the lone path to All-

out Conflict in our model, succinctly describes the conditions under which deterrence completely breaks down and reciprocated violence takes place. In our model, as in the real world, the unthinkable will occur when both words and deeds fail and a truly determined Defender is unable to convince an equally determined Challenger it intends to resist every step of the way. Both world wars, unfortunately, are harsh testimony that this dire possibility can indeed occur.

6. Summary and Conclusion

It is easy to see classical deterrence theory and the spiral model as polar opposites. Deterrence theory argues that carefully calibrated threats, judiciously applied, can stabilize a status quo and prevent deadly conflicts from developing or intensifying. Spiral theorists, on the other hand, worry that this plan for deterrence is really a prescription for disaster. Making much of the analogy with the sequence of events prior to the First World War, they claim that threats lead only to counter-threats, and that threats are inevitably reciprocated and escalated to the point that violence is unavoidable.

In this article we use a generic model of extended deterrence and escalation to evaluate the claims of both camps. Analyzing the *Asymmetric Escalation Game* with incomplete information permits us to determine when traditional deterrence succeeds, when conflicts occur but remain limited, and when they are likely to spiral out of control.

One player in this game, Challenger, chooses either to accept the status quo or to initiate a lower-level conflict. The opponent, Defender, threatens to resist a limited challenge by responding-in-kind. If there is already a limited conflict, either player can escalate it; both players threaten to counter-escalate if the other escalates first. Defender, therefore, has two deterrent threats, at the lower and higher levels of conflict, while Challenger has one, at the higher level. A player is called Hard if it prefers to execute a threat, and Soft otherwise. Thus there are two types of Challenger, Hard and Soft, and four types of Defender, Hard/Hard, Hard/Soft, Soft/Hard and Soft/Soft. The credibility of a threat refers to the probability the player is Hard.

For the purposes of this analysis, we presume that Defender believes it likely that Challenger is Hard, i.e. would prefer to counter-escalate, rather than give in, should Defender escalate first. We believe that this is the more appropriate context for a comparison of deterrence and spiral theory. *Ceteris paribus*, the status quo is more stable, and conflict spirals less probable, when Defender sees Challenger as likely Soft, i.e. unwilling to enter a high-level conflict.

Our analysis of deterrence and the spiral theory is based on identification of all the Perfect Bayesian Equilibria of the Asymmetric Escalation Game with incomplete information. These equilibria can be thought of as the predictions of the model – which specific behavioral patterns can occur under rational play, and in what circumstances. In other words, we try to associate particular Perfect Bayesian Equilibria, and the specific circumstances that give rise to them, to successful deterrence, limited conflicts, and escalation spirals.

As it turns out, one family of Perfect Bayesian Equilibria requires that the players hold beliefs that are especially implausible when Challenger is likely Hard. We dismiss this family, and that leaves just five possible equilibria. Furthermore, only two of these can exist at once, and one of them is always the No-Response Equilibrium.

Rational play under the No-Response Equilibrium is easy to describe. Without regard to credibilities or beliefs about them, Challenger always initiates and Defender always capitulates. Deterrence fails, in the traditional sense, but there is never any escalation. This behavior pattern lies outside the purview of both deterrence and spiral theory, except insofar as it shows Defender's response being deterred by Challenger. In the Asymmetric Escalation Game with incomplete information, a No-Response Equilibrium is always a rational possibility, but it throws no light at all on either deterrence or escalation spirals.

We call the remaining four Perfect Bayesian Equilibria the Spiral Family. Exactly one member of this family always co-exists with the No-Response Equilibrium – which one, is determined by Defender's credibility parameters. Within the Spiral Family, two equilibria are easy to identify with successful deterrence, one with limited conflict, and one with a conflict spiral. These equilibria are mutually exclusive, so knowing when they occur in the model enables us to formulate a prediction of when each of these behavior patterns is likely to be observed.

In our model, traditional deterrence – which we associate with the certain preservation of the status quo – is possible, provided that *both* of Defender's threats are credible enough to dissuade Challenger from a challenge. For deterrence to succeed, Defender must convince Challenger that it is likely prepared to endure an all-out (strategic) conflict, and also that it is likely willing to respond at the lower (tactical) level. In other words, deployment policies like Massive Retaliation that depend primarily on escalatory threats to deter initial challenges are not well-suited to deterring a Challenger who would likely endure a strategic conflict rather than give in (Zagare and Kilgour, 1993b). In part, the reason is straightforward: because Challenger's threat is highly credible, *Defender* will tend to be deterred from escalating first. It is hardly surprising, therefore, that Defender's threat to respond-in-kind is also critical, not only for establish-

ing traditional deterrence, but also for determining the dynamics of play under the remaining two equilibria.

When the credibilities of Defender's first- and second-level threats fall too low, Challenger initiates. If Defender is unwilling to endure a fight (and this is often true, as Defender's credibilities are low), then it capitulates and the game ends, no matter which of the remaining two equilibria happens to be in play. Nonetheless, other behavior patterns are possible in the event that Defender is willing to fight at one or both levels. After Defender responds, deterrence might be re-established, and conflict contained, at the tactical level; another possibility, though, is escalation to the strategic level.

Once again, Defender's credibilities are the key determinant. After it observes an unexpected response-in-kind, Challenger revises its original estimates of Defender's type. For a conflict to be limited, Challenger must conclude, having observed Defender to be tactically Hard, that it is likely strategically Hard as well. In other words, our model indicates that the crucial variable is the conditional probability that Defender is strategically Hard, given that it is tactically Hard. This probability, called $p_{\text{Str|Tac}}$ in Table 2, represents Challenger's revised belief that Defender is of type HH, after Defender has responded-in-kind. If Challenger finds it sufficiently probable that Defender is strategically Hard, then conflict is capped at the tactical level. If not, there is an escalation spiral.

In one sense, then, our model indicates that both limited conflicts and conflict spirals depend on unanticipated events. Acute crises and limited conflicts are largely accidental by-products of interstate competition. Further, confirming the suspicions of most spiral theorists, many all-out conflicts are situations that states blunder into, each one anticipating that it can out-escalate the other. They are truly events that no one wants.[19]

The Korean War is a clear example of a conflict that was capped when a second-level threat suddenly gained high credibility. According to de Rivera (1968: 53), after UN forces crossed the 38th parallel in 1950, 'the Assistant Secretary of State for Far Eastern Affairs [like other senior US officials] did not expect [a Chinese] invasion and, hence, failed to detect it even when he was confronted with a rather strong signal' (also see Lampton, 1973: 28). But soon after this unexpected event occurred, the UN command adjusted its actions. Fearing a wider war with the Chinese and perhaps the Soviets, US Secretary of Defense George Marshall decided to 'use all available political, economic and psychological action to limit the war' (quoted in Gacek, 1994: 57).

19. Of course, this does not apply to all wars. For instance, Prussia wanted war with Austria in 1866. Clearly, Austria's deterrent threat lacked *capability* in the sense defined earlier (see footnote 3, p. 62).

Both World Wars illustrate the second behavioral pattern: conflict that spirals to the highest level after unanticipated resistance. The difference is that in this case Challenger incorrectly believes that any resistance by Defender will be token. Escalation, therefore, is seen as a way to coerce Defender into submission.

In the years prior to the First World War, for instance, Great Britain chose not to conscript and maintain a large standing army, thereby limiting its ability to defend its continental allies. Rather, the British relied primarily on an escalatory threat (i.e. its fleet) to deter war. Germany knew all this, but found the escalatory threat alone to be insufficiently credible.[20]

> Up to the last prewar days [British Foreign Minister Sir Edward] Grey was discussing with the Germans what it would take to keep Britain neutral; the majority of the Cabinet regarded it as possible not to come to the aid of the French; many thought that Britain need not go to war if Belgium was invaded; and even after the idea of war was accepted, many thought Britain should not send an army to the Continent. Not only could Britain's friends and enemies not be sure what the British would do until the last minute, the British themselves did not know. In those circumstances it may not be surprising that even so cautious and conservative a man as [German Chancellor] Bethmann [Holweg] was willing to take the great risk that brought on the war. (Kagan, 1995: 211)

Much the same could be said about the backdrop to the Second World War. In attempting to appease the Germans, the British and French only encouraged aggression. In the end, Hitler came to believe that events like the invasion of Poland would not provoke the British into fighting. Unfortunately, like Bethmann Holweg before him, he was wrong.

Deterrence theorists and proponents of the spiral model have drawn different lessons from these and similar events, leading Jervis (1976: 84) to remark that these two conceptual models 'contradict each other at every point'. Classical deterrence theorists like Kagan claim that war follows when real or intended threats are not convincingly communicated. Spiral theorists argue that wars are rooted in the threats implied by an accelerating arms race, by a military alliance, or by a standing army. Empirically, Jervis finds that:

> neither theory is confirmed all the time. There are lots of cases in which arms have been increased, aggression deterred, significant gains made, without setting off spirals. And there are also many instances in which the use of power and force has not only failed or even left the state worse off than it was originally ... but has led to mutual insecurity and misunderstandings that harmed both sides.

20. The architect of Germany's war plan, Count Alfred von Schlieffen, thought the British *would* intervene in a continental war. For Von Schlieffen in 1906, then, Britain's first-level threat *was* credible. Unfortunately, while credible, it was not capable. Both Schlieffen and his successors discounted the military impact of Britain's small expeditionary force (Kagan, 1995: 212).

Our analysis helps to explain both why and when. Successful deterrence and conflict spirals are events that take place under very different sets of circumstances. Deterrence theory and the spiral model are complements rather than substitutes. Each is inspired by a distinct theoretical and empirical dynamic. Classical deterrence theorists are quite right to assert that capable and credible threats have the potential to avert disaster and prevent conflict. And spiral theorists are equally correct in pointing out that misjudgments and unrealistic expectations are but a prelude to catastrophe.[21] Our conclusion is that empirical attempts to validate either theoretical framework at the expense of the other are doomed to failure. The dichotomy posed by Jervis is false. The real world is more complex and more varied than either deterrence or spiral theorists admit. We believe that our model and our analysis have captured fundamental features of the complexity and the variety of real-world interstate interactions.

REFERENCES

Blainey, Geoffrey (1988) *The Causes of War*, 3rd edn. New York: Free Press.

Carlson, Lisa J. (1995) 'A Theory of Escalation and International Conflict', *Journal of Conflict Resolution* 39: 511–34.

Danilovic, Vesna (1995) 'Major Powers, Crisis Escalation, and War', PhD dissertation, State University of New York at Buffalo.

de Rivera, Joseph (1968) *The Psychological Dimension of Foreign Policy*. Columbus, OH: Merrill.

Fudenberg, Drew and Jean Tirole (1991) *Game Theory*. Cambridge, MA: MIT Press.

Gacek, Christopher M. (1994) *The Logic of Force: The Dilemma of Limited War in American Foreign Policy*. New York: Columbia University Press.

Hobbes, Thomas (1968 [1651]) *Leviathan*, ed. C. B. Macpherson. Harmondsworth: Penguin.

Holsti, Ole R., Robert C. North and Richard A. Brody (1968) 'Perception and Action in the 1914 Crisis', in J. David Singer (ed.) *Quantitative International Politics: Insights and Evidence*. New York: Free Press.

Jervis, Robert (1976) *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press.

Jervis, Robert (1978) 'Cooperation under the Security Dilemma', *World Politics* 30: 167–214.

Kagan, Donald (1995) *On the Origins of War and the Preservation of Peace*. New York: Doubleday.

Kaufmann, William (1956) 'The Requirements of Deterrence', in William Kaufmann (ed.) *Military Policy and National Security*. Princeton, NJ: Princeton University Press.

Kilgour, D. Marc and Frank C. Zagare (1991) 'Credibility, Uncertainty, and Deterrence', *American Journal of Political Science* 35: 305–34.

21. Parenthetically we note that these conclusions are entirely consistent with power transition theory which argues that parity is a necessary but not a sufficient condition for major power war (Organski and Kugler, 1980). By shedding light on the conditions associated with successful deterrence and with conflict spirals, our findings refine and extend power transition's explanatory power. For an explicit attempt to explore the connection between power shift theories like power transition and strategic uncertainty, see Powell (1996) and the literature cited therein.

Kilgour, D. Marc and Frank C. Zagare (1994) 'Using Game Theory To Analyze a General Two-level Escalation Model', Paper delivered at the Annual Meeting of the American Political Science Association, New York, NY, 1–4 September.

Kydd, Andrew (1997) 'Game Theory and the Spiral Model', *World Politics* 49: 371–400.

Lampton, David M. (1973) 'The U.S. Image of Peking in Three International Crises', *The Western Political Quarterly* 26: 28–50.

Lebow, Richard Ned (1984) 'Windows of Opportunity: Do States Jump Through Them?', *International Security* 9: 147–86.

Massie, Robert K. (1991) *Dreadnought: Britain, Germany, and the Coming of the Great War*. New York: Ballantine Books.

Morrow, James D. (1989a) 'Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining', *American Journal of Political Science* 33: 941–72.

Morrow, James D. (1989b) 'Bargaining in Repeated Crises: A Limited Information Model.' In Peter C. Ordeshook (ed.) *Models of Strategic Choice in Politics*. Ann Arbor: University of Michigan Press.

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

O'Neill, Barry (1994) 'Game Theory Models on Peace and War', in Robert J. Aumann and Sergui Hart (eds) *Handbook of Game Theory with Economic Applications*, Vol. 2. Amsterdam: Elsevier.

Organski, A. F. K. and Jacek Kugler (1980) *The War Ledger*. Chicago: University of Chicago Press.

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

Powell, Robert (1996) 'Uncertainty, Shifting Power, and Appeasement, *American Political Science Review* 90: 749–64.

Pruitt, Dean G. (1969) 'Stability and Sudden Change in Interpersonal and International Affairs', *Journal of Conflict Resolution* 13: 392–408.

Richardson, Lewis F. (1960) *Arms and Insecurity: A Mathematical Study of the Causes and Origins of War*. Pittsburgh: Boxwood Press.

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

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

Selten, Reinhard (1975) 'A Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games', *International Journal of Game Theory* 4: 25–55.

Smoke, Richard (1977) *War: Controlling Escalation*. Cambridge, MA: Harvard University Press.

Snyder, Glenn H. (1961) *Deterrence and Defense*. Princeton, NJ: Princeton University Press.

Zagare, Frank C. (1987) *The Dynamics of Deterrence*. Chicago: University of Chicago Press.

Zagare, Frank C. (1990a) 'Rationality and Deterrence', *World Politics* 42: 238–60.

Zagare, Frank C. (1990b) 'The Dynamics of Escalation', *Information and Decision Technologies* 16: 249–61.

Zagare, Frank C. (1992) 'NATO, Rational Escalation and Flexible Response', *Journal of Peace Research* 29: 435–54.

Zagare, Frank C. (1996) 'Classical Deterrence Theory: A Critical Assessment', *International Interactions* 21: 365–87.

Zagare, Frank C. and D. Marc Kilgour (1993a) 'Asymmetric Deterrence', *International Studies Quarterly* 37: 1–27.

Zagare, Frank C. and D. Marc Kilgour (1993b) 'Modeling Massive Retaliation', *Conflict Management and Peace Science* 13(1) (Spring): 61–86.

Zagare, Frank C. and D. Marc Kilgour (1995) 'Assessing Competing Defense Postures: The Strategic Implications of "Flexible Response"', *World Politics* 47: 373–417.

Zagare, Frank C. and D. Marc Kilgour (1996) 'Limited War, Crisis Escalation and Extended Deterrence', Mimeo, State University of New York at Buffalo and Wilfrid Laurier University.

FRANK C. ZAGARE is Professor and Chair, Department of Political Science, State University of New York at Buffalo. He is author of *The Dynamics of Deterrence* (Chicago, 1987) and *Game Theory: Concepts and Applications* (Sage, 1984), editor of *Modeling International Conflict* (Gordon and Breach, 1990), and co-editor of *Exploring the Stability of Deterrence* (Lynne Rienner, 1987). His work has appeared in *American Journal of Political Science*, *World Politics*, *International Studies Quarterly*, *Journal of Peace Research*, *Conflict Management and Peace Science*, *Theory and Decision*, *Journal of Conflict Resolution*, *International Interactions*, and *Synthese*. Professor Zagare's current research involves the reformulation of classical deterrence theory and the application of game-theoretic models to international affairs. ADDRESS: Department of Political Science, State University of New York at Buffalo, 520 Park Hall, Buffalo, NY 14260 USA. [email: fczagare@acsu.buffalo.edu]

MARC KILGOUR is Professor of Mathematics at Wilfrid Laurier University in Waterloo, Ontario, Director of the Laurier Centre for Military Strategic and Disarmament Studies, and Adjunct Professor of Systems Design Engineering at the University of Waterloo. His primary research interests lie in the application of game theory and related formal techniques to questions related to decision-making in a variety of contexts. He has published widely in many fields, including mathematics, operations research, political science, international security, systems engineering, biology, economics, and environmental management. ADDRESS: Laurier Centre for Military Strategic and Disarmament Studies, Wilfrid Laurier University, Waterloo, Ontario N2L 3C5 Canada. [email: mkilgour@machl.wlu.ca]

Paper submitted 2 January 1997; accepted for publication 29 July 1997.