The dynamics of escalation

F.C. ZAGARE
Department of Political Science, State University of New York at Buffalo, Buffalo, NY 14260, U.S.A.

Received June 1989
Revised August 1989
Communicated by K.W. Hipel

In this paper I use a theory of moves framework to explore the dynamics of two-stage deterrence games. The present study differs from other applications of this framework to national security questions in that the structural characteristics of two distinct deterrence games are linked by way of a common outcome. It is through this linkage process that the dynamics of the escalatory process are analyzed. Several interesting insights into the nature of these games were discovered. First, it was found that deterrence is stable, and escalation is not rational, as long as *neither* player in the second game possesses a credible retaliatory threat. Interestingly, no such pattern exists when each player's threat in the second stage is credible. It was also discovered that escalation dominance confers a distinct advantage upon a player. In all cases the player with escalation dominance 'wins' the game.

1. Introduction

In recent years the study of interstate conflict has benefited from a number of technical advances in the field of game theory.[1] Most important have been the refinements to the notion of an equilibrium outcome (see, inter alia, [5, 6, 11, 12, 14, 21]), resulting not only in more dynamic models and applications but also a closer correspondence between these theoretical structures and the real world of international politics. The purpose of this paper is to explicate the logic of one of these analytic techniques (i.e. the *theory of moves*) and to use it to explore the dynamics of the escalation process.

Escalation is a phenomenon common to a wide spectrum of decision-making situations, ranging from family squabbles to corporate price wars. In international politics, the dynamic underlying this process is of particular relevance to specialists in the areas of national security and crisis decision-making.[2] Most strategic analysts are of the opinion that the next world war, should one occur, will not be the result of a 'bolt from the blue'; rather the gravest threat to international stability is generally taken to be an escalation of some minor incident or conflict to a major international crisis and, ultimately, to a global thermonuclear war.

2. General deterrence and the credibility of threats

To explore the dynamics of escalation, I begin by positing a generalized situation of mutual deterrence (see Fig. 1) in which each of two adversarial states, Nation A and Nation B, is trying to prevent the other from upsetting the status quo.[3] To simplify matters, assume that each side has two broad strategic choices, either to cooperate (C) with the other by supporting the status quo, or not to cooperate (D) by moving to overturn it. These choices lead to four equally broad outcomes: if each state cooperates, the status quo (CC) persists; if one state cooperates, and the other does not [either (CD) or (DC)], the state which does not cooperate gains an advantage; and if neither side cooperates, conflict (DD) is implied.

Given any situation of *mutual* deterrence, cer-

[1] For a summary of recent developments, see Roth [19] and Zagare [25].

[2] The first serious academic study of the connection between escalation and war is Smoke's [22]. For a recent review of the literature, see Freedman [7].

[3] Since a 'game' is defined, inter alia, by the rules which govern the strategy choices of the players, I have deliberately avoided referring to the structures I describe in this section as 'games'. In subsequent sections, I posit two different sets of rules and thereby transform the various situations of mutual deterrence into game forms. Nevertheless, to facilitate the subsequent exposition, I shall use as labels for these deterrence situations names normally attached to those 2×2 games with similar structural characteristics.

North-Holland
Information and Decision Technologies 16 (1990) 249–261

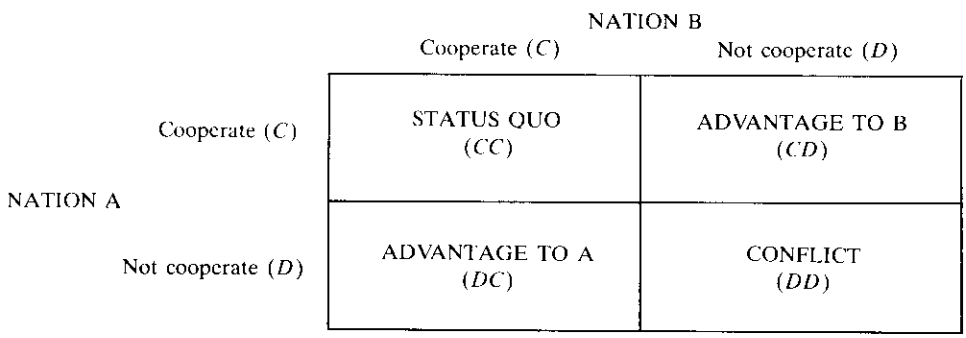

Fig. 1. A generalized deterrence situation.

tain preference relationships are implied. For example, it seems safe to assume that each player prefers the status quo to the outcome associated with the other's advantage; if this assumption were not made, neither player would be interested in deterring the other. Furthermore, it follows that each player prefers the outcome associated with its own advantage to the status quo; otherwise there would be no need for the other player to deter the first. Thus, by definition, a relationship of mutual deterrence will satisfy the following restrictions on the preferences of the two players:

$$\text{for Nation A: } (DC) > (CC) > (CD)\,, \qquad (1a)$$

and

$$\text{for Nation B: } (CD) > (CC) > (DC)\,, \qquad (1b)$$

where '>' means 'is preferred to'.

In order to completely define a deterrence situation, the relationship between these three outcomes and the final outcome, (DD), must be specified. To simplify the exposition, assume that each nation prefers the status quo (CC) to mutual punishment (DD). This, too, seems like a safe assumption; without it, neither player possesses a deterrent threat and, as demonstrated in Zagare [26], deterrence is not possible. With this simplification, only three situations of mutual deterrence are possible (see Fig. 2): 'Prisoners' Dilemma' [Fig. 2(a)], 'Called Bluff' [Fig. 2(b)], and 'Chicken' [Fig. 2(c)].

Each cell of these figures contains an ordered pair. By convention, the first entry represents the payoff of the row player (here, Nation A) and the second, the payoff of the column player (here, Nation B). In Fig. 2, these outcomes are ranked from '1' to '4' with '4' representing each player's best outcome; '3' each player's next-best outcome, and so on.

In Prisoners' Dilemma, each player prefers mutual punishment to the outcome associated with the other's advantage; in Called Bluff, only one player (i.e. Nation A) exhibits this preference pattern; and in Chicken, both players prefer the other's advantage to mutual punishment and conflict.

One way to think about the differences among these structures is in terms of the *credibility* of each player's deterrent threat. The outcome associated with mutual punishment, (DD), represents the threat upon which the deterrence relationship rests. In order to deter the other player from upsetting the status quo, each player threatens to retaliate by counter-moving from the outcome associated with the other's advantage [either (CD) or (DC)] to mutual punishment (DD).

Deterrent threats can be either credible or incredible. As discussed in detail by Kilgour and Zagare [13], strategic analysts generally take a credible threat to be synonymous with a threat which is believed. Believability, in turn, is normally equated with rationality. Credible threats, therefore, are believable threats; believable threats are threats which are rational to carry out. Thus, only rational threats are credible [6, 21, 26].

In the literature of decision theory, a rational

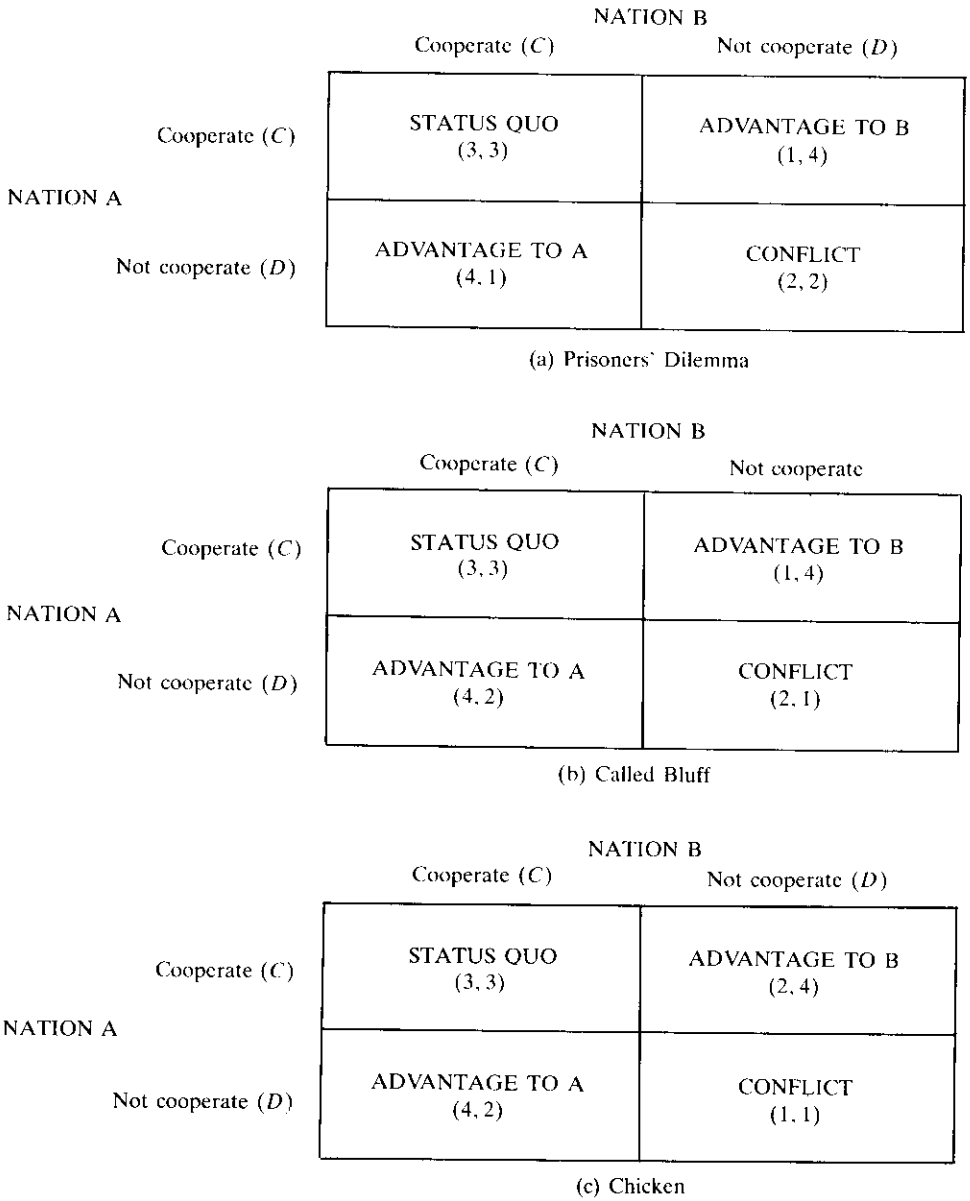

Fig. 2. Three Mutual Deterrence situations.

choice is minimally defined to be a choice which is consistent with a player's preferences [15]. Thus, because the situations of Fig. 2 differ only with respect to the preferences of the players between executing the threat and accepting the outcome associated with the opponent's advantage, they can also be viewed as situations which are distinguished by differences in the inherent credibility of each player's threat. In the first situation (Prisoners' Dilemma), since each player prefers mutual punishment to his opponent's advantage, each has a credible (i.e. rational) retaliatory threat. In the second (Called Bluff), since only A has this preference, A's threat is credible and B's is not. And in the last situation (Chicken), neither player's threat is credible since each prefers capitulation to mutual punishment.

3. Mutual deterrence: A theory of moves analysis

What is the connection between deterrence stability and threat credibility? To answer this question, I now employ a dynamic game-theoretic framework called the 'theory of moves'. As developed by Brams and Wittman [5] and extended by others [2, 11, 12, 26], the theory of moves 'describes optimal strategic calculations in normal-form games in which the players can move and countermove from an initial outcome in sequential play' [2, p. 184].

At the heart of the theory of moves framework is the concept of a *nonmyopic equilibrium*. Underlying this equilibrium concept is the assumption that, after an initial strategy choice, each player in a sequential game can make conditional *and* sequential moves from the initial outcome or status quo point, and is able to evaluate the long-term consequences of such a departure. More specifically, the concept of a nonmyopic equilibrium assumes that the following rules of play operate in a 2×2 ordinal game:

(1) Both players simultaneously choose strategies, thereby defining an *initial outcome* of the game or, alternately, in the interpretation used in this paper, an initial outcome (or status quo) is imposed on the players by circumstances.

(2) Once at an initial outcome, either player can unilaterally switch his strategy and change the outcome to a *subsequent outcome*.

(3) The other player can respond by unilaterally switching his strategy, thereby changing the subsequent outcome to a new subsequent outcome.

(4) These strictly alternating moves continue until the player with the next move chooses not to switch his strategy. When this happens, the game terminates, and the outcome reached is the *final outcome* [3].

These rules bear more than a passing resemblance to the conditions typically present in deterrence games. As Wagner [24, p. 342] notes, 'a feature common to all situations involving the use of military force is that one government does something, whereupon another either does or does not do something in response, whereupon the first in turn either does or does not reply to the second's response, and so forth'. Thus, the dynamic choice framework associated with the theory of moves renders it an extremely attractive methodology for assessing the conditions associated with deterrence stability, or the lack thereof.

Given these rules, and the ability of the players to calculate the consequences of a departure from an initial outcome, two conditions must be met for an initial outcome to be considered a nonmyopic equilibrium. First, neither player must perceive an advantage in departing from it, and second, there must be *termination* of the move, counter-move sequence, that is, the sequence of moves and counter-moves must not cycle back to the initial outcome. Brams and Wittman [5] assume that there will be termination if an outcome is reached in the sequential move process whereby the player with the next move can ensure his best outcome by staying at it.[4]

The concept of a nonmyopic equilibrium, then, is a look-ahead idea that assumes that a player will evaluate the long-term consequences of departing from an initial outcome, taking into account the probable response of the other player, his own counter-response, subsequent counter-responses, and so on. If, for *both* players, the starting outcome is preferred to the outcome each player calculates he will end up at by making an initial departure, the starting outcome is a nonmyopic equilibrium.

To determine whether deterrence is stable in each of the three situations of Fig. 2, one simply tests for the nonmyopic stability of the status quo outcome in each one. To do this, a backward induction process is applied to a tree such as is given by Fig. 3. This particular tree lists the sequence of moves and countermoves away from the status quo in the game of Fig. 2(a) (Prisoners' Dilemma), given an initial departure from this outcome by the row player, Nation A. It is easy to demonstrate that, *under these conditions*, deterrence constitutes a stable relationship when each player has a credible retaliatory threat.

To see this, consider first Nation B's choice at the last node of this tree. As the arrow indicates, at this node, B would rationally choose to stay at his best outcome, (1, 4), rather than move to his

[4]For a generalization of this concept where the termination requirement is dropped, see Kilgour [11, 12].

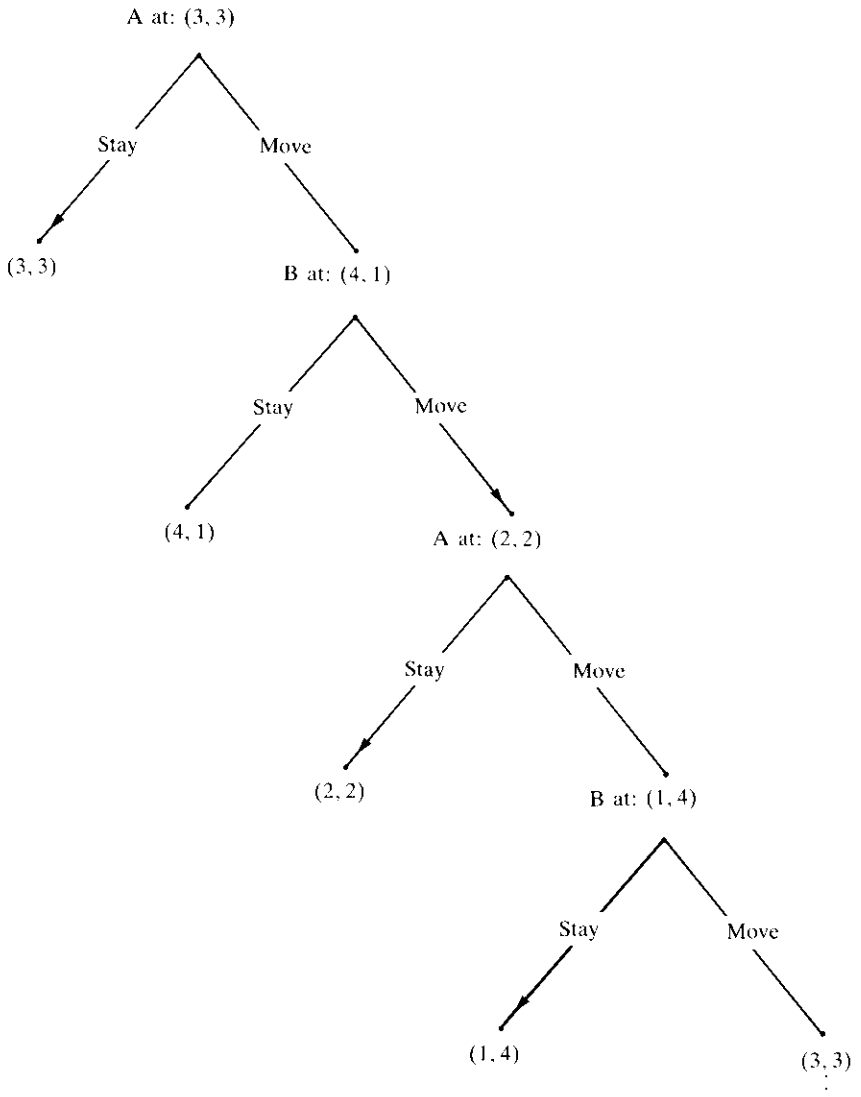

Fig. 3. Game tree representation of moves in Prisoners' Dilemma, starting with A at (3, 3). *Legend*: → indicates rational choice.

next-best outcome, (3, 3).[5] Moreover, given that B would stay at (1, 4), it is also clear that A should, at the previous node, stay at (2, 2), his next-worse outcome, rather than permit B to end the sequential game at his worst (and B's best) outcome (1, 4). Similarly, at the previous node (4, 1), B would rationally move away from his worst outcome in order to induce his next-worst outcome at (2, 2), which would become the final outcome in this sequential game because of A's incentive to terminate the sequential move process there. Finally, at the top of the tree, A would rationally stay at (3, 3), his next-best outcome, rather than precipitate a sequence of moves and countermoves which would rationally terminate at his next-worst outcome (2, 2). Thus, A has no incentive to upset the status quo in the game of Fig. 3.

By symmetry, neither does B. Since neither player has a long-term incentive to move away from (3, 3), this outcome is a nonmyopic equilibrium. It is for this reason that deterrence constitutes a stable relationship between two equally

[5]Thus, the termination requirement is satisfied. This is why the tree of Fig. 3 need not be extended past this particular node.

powerful states, each with a credible and capable retaliatory threat.

Deterrence is not stable, however, when only one player has a credible threat. Applying reasoning similar to the above to the second structure of Fig. 2 (Called Bluff), it is easy to demonstrate the Nation A, the player with a credible threat, *does* have an incentive to move from the status quo. If Nation A moves from (3, 3) to (4, 2), then Nation B – the player lacking a credible threat – will not rationally choose to resist the incursion of A. The final outcome implied by A's departure from the status quo, therefore, is (4, 2). Since Nation A prefers (4, 2) to (3, 3), his incentive to depart is established. In Called Bluff, then, deterrence is unstable and the player whose threat is credible will gain an advantage.

Interestingly, deterrence can also be stable when neither player's deterrent threat is credible [see Fig. 2(c)]. Under these conditions, neither player will move to upset the status quo since the other player, moving second, will rationally counter-move to (DD) and force the first to choose between his worst and next-worst outcome. Since both Nation A and Nation B prefer the status quo to either of these other two outcomes, each lacks an incentive to upset it.

Nonetheless, as demonstrated in Zagare [26], this conclusion rests upon the supposition that a move *to*, and *through* (DD) is both logically and empirically possible. But at the strategic nuclear level, these conditions are *not* likely to be satisfied. In a nuclear crisis a decision to induce the outcome associated with mutual punishment is more likely to terminate the sequential move process, and a great deal more. Under these conditions, therefore, neither player would *rationally* make this choice. Consequently, not fearing retaliation by a rational opponent, each player has an incentive to upset the status quo in order to gain an advantage. Clearly, deterrence is not stable in a nuclear deterrence situation in which each actor's retaliatory threat lacks credibility.[6]

[6]To some extent all of the preceding conclusions can be disturbed with slightly different assumptions about the exact sequence of choices available to the players, their preferences about the status quo, their relative power and capability, and their information about the preferences of their opponent. For the details, see Kilgour and Zagare [13] and Zagare [26].

4. The dynamics of escalation

In the preceding section I illustrated the logic associated with the 'theory of moves' framework by analyzing three situations of mutual deterrence. In this section I shall extend this framework to explore the dynamics of escalation.[7] For the purpose of determining the conditions which lead to stable deterrence in a crisis situation, I posit a multi-level game in which the decisions made at a beginning stage determine whether that stage will be terminal, or merely a point through which the play of the game proceeds. This view of escalation is consistent with Kahn's [9, 10] concept of an escalation ladder, the prevailing metaphor of the escalation process. Indeed, the outcome matrix summarizing the postulated decision sequence resembles a ladder, or at least a series of steps progressing upward or downward (see Fig. 4). As preliminary to a more general examination of the escalation ladder, I here explore the consequences for deterrence of those games with but two steps, controlling, as before, for variations in the credibility of each player's threat. In addition to providing an insight into the dynamics of interstate escalation, this extension of the theory of moves framework will permit a further refinement of our understanding of the connection between threat credibility and crisis stability.

4.1. Assumptions

As already noted, the underlying assumptions which define the two-stage escalation game studied here are reflected in the matrix of Fig. 4. Specifically, I now amend the rules of play postulated above and assume that:

(1) Each player begins by cooperating (C) with the other. The simultaneous choice of a (C) strategy by each player establishes a status quo outcome.

(2) Once at the status quo, either player may defect (D) from cooperation.

(3) If one player defects, the other may continue to cooperate (C), he may match the first player's defection by responding in-kind and defecting (D), or he may escalate (E) the conflict

[7]Alternative game-theoretic treatments of escalation can be found in Brams and Kilgour [4] and O'Neill [17].

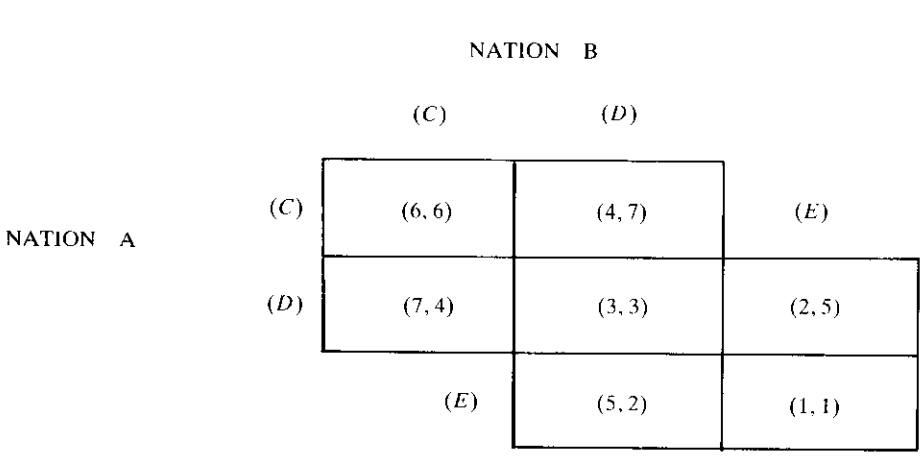

Fig. 4. Double Chicken. *Legend*: 7 = best, 6 = next-best, and so on.

by selecting a strategy qualitatively different than the level of punishment associated with a choice of (D).

(4) If the second player defects or escalates (in response to the first's defection), the first may retaliate with an escalatory move of his own.

(5) Whatever the sequence of choices, once one player escalates, the other has the option of matching this choice by also choosing to escalate.

(6)The game ends if both players initially choose to stay, or as soon as one player chooses to stay after the other defects or escalates, or once both escalate.

Notice that the four northwest and the four southeast cells of Fig. 4 share the structure of the generalized deterrence situation of Fig. 1. Fig. 4, therefore, links two distinct 2×2 deterrence situations by way of a common outcome. Specifically, the conflict outcome (DD) of the upper (northwest corner) component game also serves as the status quo outcome for the lower (southeast corner) component game. By linking the situations in this fashion, the interconnections of two very different payoff structures can be explored.

In addition to these assumptions about the nature and the sequencing of crisis decision-making, I also make several assumptions about the preferences of the players:

(7) As before, each player is assumed to prefer to defect rather than to cooperate, given that his opponent has cooperated initially. This assumption provides each player with an *immediate* incentive to move down (up?) the escalation ladder away from the status quo.

(8) Each player prefers to escalate rather than cooperate or defect, given that his opponent has already defected. I make this assumption in order to examine precisely those games with a built-in dynamic toward escalation.

(9) Nevertheless, I assume that each player prefers to gain an advantage at the lowest possible level of conflict. For example, in Fig. 4, Nation A is postulated to prefer outcome (DC) – which occurs when it defects initially and Nation B fails to retaliate – to (ED) – which results when A escalates in response to a choice of (D) by B, and B fails to retaliate by also choosing (E).

To be sure, these assumptions do not exhaust the set of reasonable postulates about the preferences of players in a two-stage game like this one. These particular assumptions, however, provide both a useful starting point and a worst-case scenario for the analysis of crisis decision-making.

Finally, I continue to assume that the players decide whether or not to defect from the status quo in full anticipation of the consequences of such a choice. In other words, the players are postulated to make nonmyopic decisions which take into account not only the immediate ramifications of their decisions, but also the long-term consequences implied by their choices.

In order to determine the rational choice of each player in a two-stage deterrence game, therefore, a tree similar to that of Fig. 3, but which reflects the added complexity inherent in the escalation process, is required. Fig. 5, which reflects assumptions (1) to (6) above, is one such tree. In this case, the tree depicts the sequence

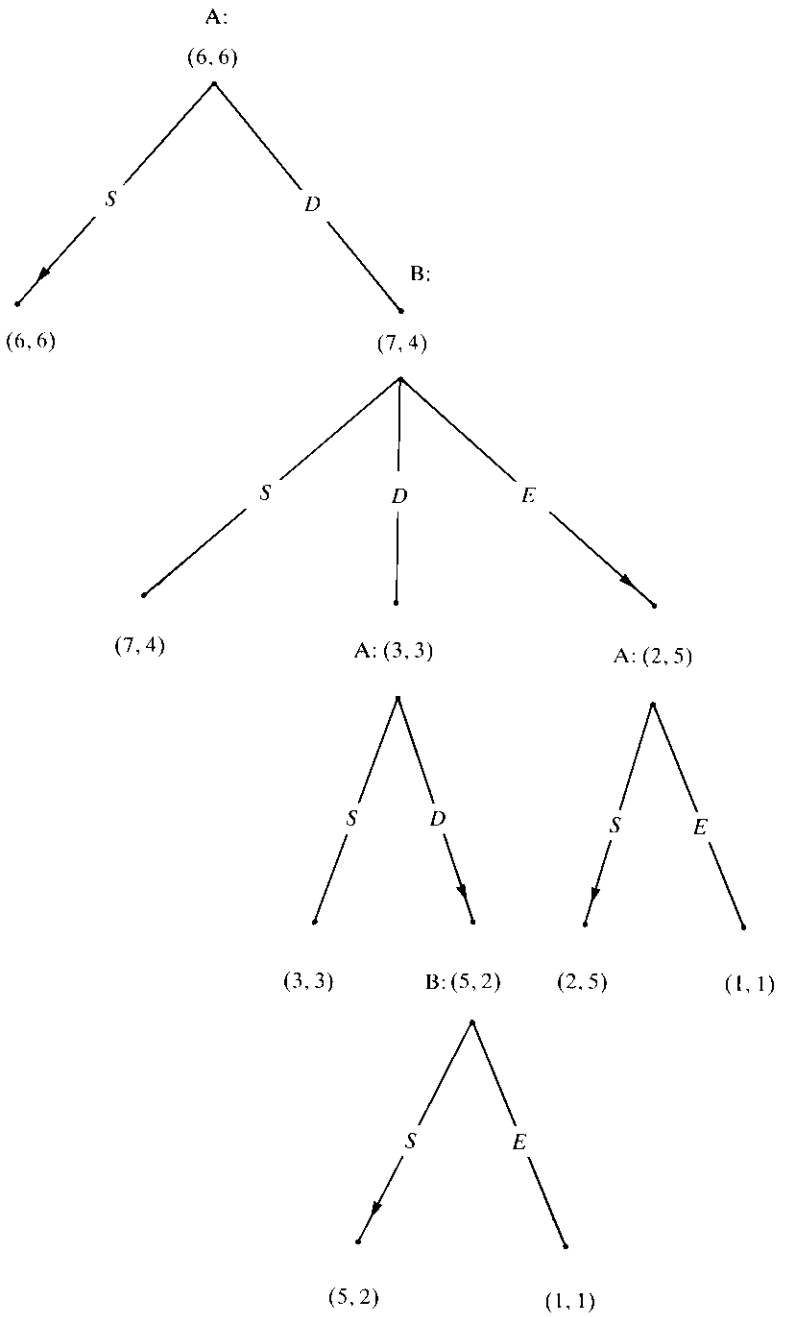

Fig. 5. Game tree representation of moves of the game of Fig. 4, starting with A at (6, 6). *Key*: S = Stay; D = Move by defecting; E = Move by escalating; ⟶ = rational choice.

of choices implied by an initial departure from the status quo by Nation A in the game given by Fig. 4. Using a backward induction process similar to that outlined above, it is easy to show that in *this* game, Nation A should choose to cooperate initially. (The rational choice of each player at each node is indicated by an arrow.) Should Nation A stay at the original status quo, his next-best outcome "6" is implied. But should he defect and move to an immediately better outcome (i.e. "7"), B will rationally escalate, thereby inducing A's next-worst outcome. At this

point, if the players are rational, the sequential move process will terminate. Since A prefers the original status quo to the outcome implied by the sequence of rational moves and counter-moves, he should stay at the status quo. By symmetry, so should Nation B. Thus, in the game of Fig. 4, deterrence is stable and escalation is not rational.

4.2. *Some illustrative games*

The two-stage deterrence game of Fig. 4 is interesting in its own right. Notice that in the 2×2 component game comprising the northwest section of Fig. 4, each player prefers to capitulate should his opponent defect. Thus, this component shares the structural characteristics of Chicken [see Fig. 2(c)]. The same is also true of the component in the southeast section of Fig. 4. In this region of the matrix, neither player prefers to match an escalatory move by the other; rather, each prefers to accept the outcome associated with an advantage for his opponent (i.e., either (DE) or (ED)). Thus, the structure of Fig. 4 links, by way of a common outcome, two games of Chicken. Or, put in a slightly different way, in the game depicted by Fig. 4, each player lacks a credible retaliatory threat at each stage of the game.

That deterrence is stable in the game of Fig. 4 is quite remarkable. Recall that deterrence is not stable in Chicken when the conflict outcome is postulated to be a terminal outcome. But in the two-stage game linking component games with the structural characteristics of Chicken, the instability disappears. Intuitively, the reason for this is straightforward. Each player is deterred from upsetting the status quo in the first stage because he prefers to avoid an end game which, because of the postulated sequence of moves and player preferences, favors his opponent. Thus, deterrence can be stable in a linked game even when stability is not characteristic of its constituent parts. Surprisingly, then, in a two-stage game in which neither player has a credible threat at either stage, deterrence *is* stable.

Significantly, the stability exhibited by the game of Fig. 4 evaporates when the threat of each player in the second stage of the game is assumed to be credible. In the game of Fig. 6, wherein neither player has a credible threat at the initial stage but where each player's threat to escalate at the subsequent stage is indeed credible, deterrence is *not* stable. In fact, each player has an incentive to move from the status quo although neither has an incentive to retaliate should the other move first. In this game, then, deterrence is unlikely and the outcome is uncertain.

The dynamics of this game illustrate what Snyder [23] calls the 'stability–instability' paradox, namely the fact that the stability of an end game (i.e. Prisoners' Dilemma) may promote instability at a lower level of conflict. The instability of the game of Fig. 6 is all the more startling given its juxtaposition with the game of Fig. 4. The only difference between these two structures is the presence of mutually credible threats in the latter case, but not in the former. Thus, the lack of stability in the game of Fig. 6

NATION A \ NATION B	(C)	(D)	(E)
(C)	(6, 6)	(4, 7)	
(D)	(7, 4)	(3, 3)	(1, 5)
(E)		(5, 1)	(2, 2)

Fig. 6. Chicken – Prisoners' Dilemma. *Legend*: 7 = best, 6 = next-best, and so on.

can be directly attributed to the credibility of each player's deterrent threat in the end game! Mutually credible threats, therefore, may be a double-edged sword, promoting stability in some cases and destroying it in others.

Given the above, it is quite plausible to suggest that one reason for the series of superpower crises over Berlin during the late 1950s and the early 1960s is the structural similarity between the game given by Fig. 6 and the choices facing Soviet and American decision-makers at the time. As Quester [18] notes, one problem associated with the Eisenhower administration's policy of Massive Retaliation and its almost total reliance on nuclear deterrence was the fact that the nuclear threat 'might indeed be credible all-around'. Thus, in a setting 'where any military initiative had to fall to the West rather than the Soviet bloc', Western leaders might be presented with a very unpalatable choice: either nuclear war or capitulation. In other words, given mutually credible escalatory threats, each side would be deterred at the strategic level but, somewhat paradoxically, would not be deterred from seeking unilateral advantages at the margin.

But what if only one player has a credible threat in the end game, thereby creating an asymmetry of motivation? Fig. 7 depicts one game exhibiting this characteristic. In this example, each player's threat is assumed to be credible in the first stage (Prisoners' Dilemma), but only Nation A's threat is rational in the second game (Called Bluff). As the arrow in Fig. 7 indicates, deterrence is not stable under these conditions and the player with the credible threat at the highest level (i.e. A) should gain the advantage. Illustrated here is the dynamic implicit in the notion of *escalation dominance* [7, 10] which I define to be an asymmetry of credibility in the final stage of a two-stage game, or in the first stage when both players have credible threats in the final stage of a game.[8] Clearly, asymmetries in a deterrence relationship, such as the one exhibited by the game of Fig. 7, are potential destabilizing forces in interstate politics. Precisely because one state might be motivated to exploit such an asymmetry, each state in a mutual deterrent situation should fear perceived or real inequalities of strategic capability. As early as 1959, Morgenstern [16] recognized that it was in the interest of the United States for the Soviet Union to possess an invulnerable second-strike force, and vice versa. And today, it is precisely the instability imputed to games like Fig. 7 that lies behind the objection of many strategic thinkers to the Strategic Defense Initiative under development in the United States.

[8]Kahn [10, p. 290] defines escalation dominance to be 'a capacity, other things being equal, to enable the side possessing it to enjoy marked advantages in a given region of the escalation ladder... It depends on the net effect of the competing capabilities on the rung being occupied, the estimate by each side of what would happen if the confrontation moved to these other rungs, and the means each side has to shift the confrontation to these other rungs'.

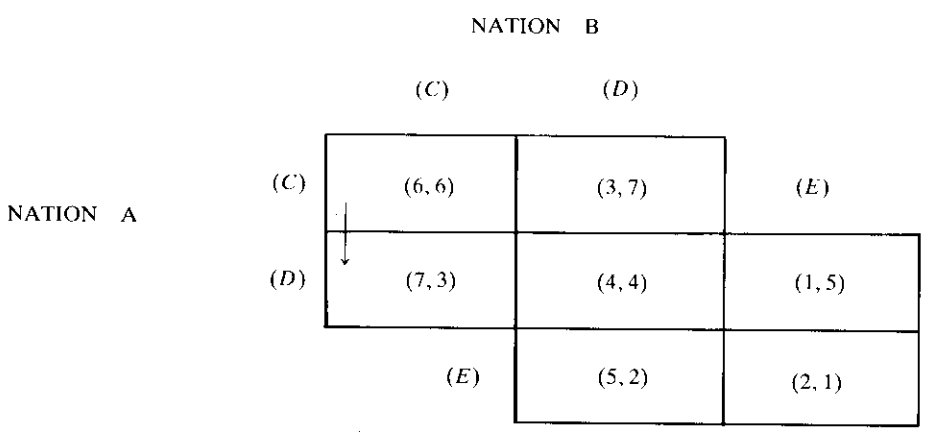

Fig. 7. Prisoners' Dilemma – Called Bluff. *Legend*: 7 = best, 6 = next-best, and so on.

5. Results

Given the assumptions (1)–(9) noted above, there are precisely 10 different two-stage deterrence games. These game – which are listed in Table 1 – are distinguished only by different assumptions about the credibility of each nation's deterrent threat at each stage of the game. Additional information concerning the outcome implied by a theory of moves analysis is also contained in this table.

A number of interesting insights into the dynamics of escalation can be gleaned from Table 1. First, notice that for each of the three two-stage games in which each player lacks a credible retaliatory threat in the second and final stage (i.e. games 1, 2 and 3), deterrence is stable and escalation is not rational. As illustrated above by the double Chicken game of Fig. 4 (game 2 in Table 1), the absence of mutually credible threats at the highest rung of the escalation ladder may actually be a stabilizing force for a deterrence relationship. Thus, deterrence stability in the nuclear age may depend less upon a fear by each player that his opponent will respond *irrationally* to an untoward action, as many deterrence theorists have speculated (see, for instance, [20, pp. 536–543]) than on the expectation by each nation that the other will respond optimally by escalating right up to the final rung of the escalation ladder. In other words, even when a player's deterrent threat lacks credibility, the ploy of feigning irratinality may not be necessary to deter an opponent in a multi-stage deterrence game similar to those modeled here.

Besides these three games, there is only one other game in which deterrence is stable and escalation irrational (game 6). In this double game of Prisoners' Dilemma, each player has a credible deterrent threat at each stage of the game. Interestingly, deterrence is not stable in the two other games (4 and 5) in which each player's threat is credible in the end game. Thus, a mutually credible threat in the final stage of the escalation ladder is neither necessary nor sufficient for deterrence stability.

There are five games in Table 1 in which one player enjoys escalation dominance. In four of these games, one player possesses a credible threat in the second stage while his opponent does not (games 7, 8, 9, and 10). In each case, the player whose threat is credible in the final stage wins, no matter what the credibility relationship is at the first stage. Thus, it is clear that escalation dominance confers an important strategic advantage on the player who possesses it.

Escalation dominance also plays a role in the resolution of game 4. Here, however, the asymmetry manifests itself in the first stage since each player's threat is credible in the second stage, but only one player's threat (i.e. A's) is credible at the first stage. Consistent with the above, it should be no surprise that deterrence is unstable and that the dominant player should win.

Finally, it is worth pointing out that in none of the postulated games does either player have an

Table 1
Two-Stage escalation games and their solutions

Game	First stage game	Second stage game	Solution	Winner
1	Called Bluff/B	Chicken	(*CC*)	Tie
2	Chicken	Chicken	(*CC*)	Tie
3	Prisoners' Dilemma	Chicken	(*CC*)	Tie
4	Called Bluff/B	Prisoners' Dilemma	(*DC*)	A
5	Chicken	Prisoners' Dilemma	(*DC*) or (*CD*)	A or B
6	Prisoners' Dilemma	Prisoners' Dilemma	(*CC*)	Tie
7	Called Bluff/B	Called Bluff/A	(*CD*)	B
8	Called Bluff/B	Called Bluff/B	(*DC*)	A
9	Chicken	Called Bluff/B	(*DC*)	A
10	Prisoner's Dilemma	Called Bluff/B	(*DC*)	A

Notes: Called Bluff/A = B has credible threat/A is bluffing.
Called Bluff/B = A has credible threat/B is bluffing.

incentive to escalate, or even to retaliate, should its opponent upset the status quo, in spite of the built-in dynamic of these games toward escalation. Put in a slightly different way, in no case is it rational for the players to move very far down the ladder. To be sure, such a result is partly a function of the complete information assumption. Given common knowledge of each other's preferences, the player with a long-term advantage can easily be identified. Thus, in *these* games, there is no need to acquire information about the depth of the other's resolve by taking tiny, though potentially destabilizing, steps down the ladder. On the other hand, this finding also requires nonmyopic players. It is perhaps this commodity, more than any other, that separates crises – like Sarajevo in 1914 – which escalate from those, like the Cuban missile crisis of 1962, which are successfully managed [1]. As Theodore Sorensen described the deliberations of President Kennedy's advisors during the missile crisis:

> We discussed what the Soviet reaction would be to any possible move by the United States, what our reaction with them would have to be to that Soviet reaction, and so on, trying to follow each of those roads to their ultimate conclusion (quoted in [8, p. 188]).

6. Summary and conclusions

In this paper I use a theory of moves framework to explore the dynamics of two-stage deterrence games. These games differ from other applications of this framework to national security questions in that the structural characteristics of two distinct deterrence situations are linked by way of a common outcome. It is through this linkage process that the dynamics of the escalatory process are analyzed.

Several interesting insights into the nature of these games were discovered. First, it was found that deterrence is stable, and escalation is not rational, as long as *neither* player in the second game possessed a credible retaliatory threat. Not only was this result unexpected, but it also stands in stark contrast to the dynamics of a one-stage game with these very characteristics.

Interestingly, no such pattern exists when each player's threat in the second stage is credible. In two cases, deterrence is unstable though escalation beyond the first stage of the game is not rational. Deterrence stability exhibits itself only when each player's threat is credible in each stage of the game.

It was also discovered that escalation dominance – defined as an asymmetry of credibility in the final stage of a two-stage game, or in the first stage when both players have credible threats in the final stage of a game – confers a distinct advantage upon a player. In all five games having this characteristic, the dominant player is advantaged.

It should also be pointed out that in no case could the players choose rationally to escalate a crisis past the first stage of a two-stage game, despite the fact that the games were purposely designed with a bias toward escalation. Moreover, this finding is invariant with respect to credibility assumptions. The slippery slope down the ladder may not be as slippery as commonly thought.

Finally, it should be noted that this analysis is intended as preliminary to a full extension of the theory of moves framework to the subject of escalation. There are many ways in which the present analysis can be extended and, indeed, some of these avenues are currently being explored. One obvious direction for expansion is the examination of deterrence games of 3, 4 and n stages. Others include controlling for variations in the capability of each player to hurt the other with an escalatory move; or in the ability of the players to control the sequential move process by forcing the other to backtrack or make the next move; or by examining the implications of the inadmissibility of certain moves. As all of these avenues have proven to be fruitful in understanding the dynamics of single-stage deterrence games [26], it will be interesting to see what these other constraints imply for the process of escalation.

Acknowledgements

I wish to thank Steven J. Brams and D. Marc Kilgour for their helpful comments on an earlier version of this paper.

References

[1] Graham T. Allison, Essence of Decision: Explaining the Cuban Missile Crisis (Little, Brown, Boston, 1971).

[2] Steven J. Brams, Superior Beings: If They Exist, How Would We Know? (Springer-Verlag, New York, 1983).

[3] Steven J. Brams, and Marek P. Hessel, Absorbing outcomes in 2×2 games, Behavioral Science 27 (1982) 393–401.

[4] Steven J. Brams and D. Marc Kilgour, Game Theory and National Security (Basil Blackwell, New York, 1988).

[5] Steven J. Brams and Donald Wittman, Nonmyopic equilibria in 2×2 games, Conflict Management and Peace Science 6 (1981) 39–62.

[6] Niall M. Fraser and Keith Hipel, Conflict Analysis: Models and Resolution (North-Holland, New York, 1984).

[7] Lawrence Freedman, On the tiger's back: The development of the concept of escalation, in: Roman Kolkowitz (Ed.), The Logic of Nuclear Terror (Allen & Unwin, Boston, 1987).

[8] Ole R. Holsti, Richard A. Brody and Robert C. North, Measuring Affect and Action in International Relations Models: Empirical Materials from the 1962 Cuban Crisis, Journal of Peace Research 1 (1964) 170–189.

[9] Herman Kahn, Thinking About The Unthinkable (Horizon Press, New York, 1962).

[10] Herman Kahn, On Escalation (Praeger, New York, 1965).

[11] D. Marc Kilgour, Equilibria for far-sighted players, Theory And Decision 16 (1984) 135–157.

[12] D. Marc Kilgour, Anticipation and stability in two-person non-cooperative games, in: Urs Luterbacher and Michael D. Ward (Eds.), Dynamic Models of International Conflict (Lynne Rienner Publishers, Boulder Co., 1985).

[13] D. Marc Kilgour and Frank C. Zagare, Uncertainty and deterrence, Unpublished, 1989.

[14] David M. Kreps and Robert Wilson, Sequential equilibria, Econometrica 50 (1982) 863–894.

[15] R. Duncan Luce and Howard Raiffa, Games and Decisions: Introduction and Critical Survey (John Wiley & Sons, New York, 1957).

[16] Oskar Morgenstern, The Question Of National Defense (Random House, New York, 1959).

[17] Barry O'Neill, International escalation and the dollar auction, Journal of Conflict Resolution 30 (1986) 33–50.

[18] George H. Quester, Nuclear Diplomacy (Dunellen, New York, 1970).

[19] Alvin E. Roth, Game-Theoretic Models Of Bargaining (Cambridge University Press, New York, 1985).

[20] Thomas C. Schelling, Arms And Influence (Yale University Press, New Haven, CT, 1966).

[21] Reinhard Selten, A re-examination of the perfectness concept for equilibrium points in extensive games, International Journal of Game Theory 4 (1975) 25–55.

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

[23] Glenn H. Snyder, The balance of power and the balance of terror, in: Paul Seabury (Ed.), Balance of Power (Chandler, San Francisco, 1965).

[24] R. Harrison Wagner, Deterrence and bargaining, Journal of Conflict Resolution 26 (1982) 329–358.

[25] Frank C. Zagare, Recent advances in game theory and political science, in: Samuel Long (Ed.), Annual Review of Political Science (Ablex Publishing Corporation, Norwood, NJ, 1986).

[26] Frank C. Zagare, The Dynamics of Deterrence (University of Chicago Press, Chicago, 1987).