

## Game Theory: Modeling Interstate Conflict

Unlike many of the approaches used in this volume to examine NATO's conflict with Serbia over Kosovo, game theory is not a theory if, by "theory," one means a set of logically interrelated propositions or hypotheses. Game theory, though, could be considered a part of the more inclusive "rational choice" paradigm, but then so could classical and structural realism, neoliberal institutionalism, or any approach that posits purposeful or goal-directed behavior.

What then is game theory? For the purposes of this chapter, game theory is best thought of as a methodology for examining strategic behavior among interacting and interdependent units. If seen as a methodology rather than as a theory, it is clear that no definitive game theoretic interpretation of a sequence of real world events such as those that led up to the conflict in Kosovo could be developed. Nonetheless, a small set of critical assumptions and concepts would set any game theoretic treatment of Kosovo apart from that of other approaches illustrated in this volume.<sup>1</sup>

In game theory, the interacting units are called *players*. Who or what constitutes a player, however, is a determination left to the individual analyst. Normally, the identification of players depends on the analyst's purpose and the specific research questions addressed. One could, for example, conceive of the conflict in Kosovo as a two-person game between Serbia and the KLA, as a two-person game between the NATO alliance and Serbia, or as a three-person game among Serbia, NATO, and the KLA. An even more detailed analysis, however, that recognized important policy differences that separated the United States from France and Germany, might specify five (or more) players. In other words, both the number and identification of players are extra-game theoretic decisions that, in principle, should be driven by theoretical and/or empirical factors rather than by methodological considerations.

A realist who was also a game theorist would likely limit the specification of players to states. But players need not be so restricted. For example, one could employ game theory from within Graham Allison's (1971) organizational process model to focus on, say, the decisionmaking process within the US government. In this game, the players—such as the White House, the Departments of State and Defense, and other important bureaucratic

departments and agencies—would be completely different *units* of analysis. But the analysis, nonetheless, could still be game theoretic.

Regardless of how they are specified, however, the players are assumed to make *choices* that, along with the choices of the other players, lead to specific *outcomes*. The specification of outcomes, like that of the players, is another important judgment call that an analyst must make. Thus a general analysis of the Kosovo conflict might simply specify “negotiation” as a possible outcome. By contrast, a more fine-grained study might distinguish between a negotiated outcome that favored the Kosovars and one that favored the Serbians.

In game theory, the players are assumed to evaluate (subjectively) the *utility* (or worth) of each possible outcome and to make choices that maximize their expected utility. Players who are expected utility maximizers are said to be instrumentally rational.<sup>2</sup> Instrumentally rational players, then, are those who *always* make choices they believe are consistent with their interests and objectives *as they define them*. In other words, instrumentally rational players are purposeful players.<sup>3</sup>

This is not to say, though, that the players evaluate the outcomes in the same way. In fact, many conflicts are traceable to disputes about the value of specific outcomes. For example, the KLA believed that by pressing its case against Serbia, an independent Kosovo would result. Serbian leader Slobodan Milosevic disagreed. Milosevic thought he could use force to stamp out the separatist movement in Kosovo and, in the process, consolidate his internal political standing.

To minimally specify a game, the players, the choices available to them, the consequences of their choices (i.e., the outcomes), and the players’ utility for each outcome must be specified. James Morrow correctly observes that “the design of the game [is] the single most important decision in modeling” (1994: 57). Note, however, that these are decisions that must be made *before* the analytic tools of game theory can be used. This means that the explanatory power of any game theoretic analysis depends less on the methodology itself than on the theoretical sensitivity of the analyst. Nonetheless, as we hope to show, game theory provides a potentially useful methodological framework for analyzing international conflicts in general, and aspects of the conflict in Kosovo in particular.

## ■ Why Game Theory?

Why use game theory to examine real world interactions? Not all game theorists agree that game theory should be used for this purpose. Morrow asserts that “individual case studies are poor tests of rational choice models.” He argues that “ex post reconstructions of historical events use information that the actors could not have at the time, subtly influencing us

away from the strategic problems they faced [and] cannot explain the case precisely because the game cannot fail to fit the specified facts of the case” (1997: 29).<sup>4</sup> Although we agree with Harry Eckstein (1975) that “critical case studies” can and do serve important theoretical purposes, Morrow is undoubtedly correct in arguing that even a theoretically informed single case study, or what Robert Bates and colleagues (1998) call an “analytic narrative,” would fall short of a *test* of a game theoretic model or, for that matter, any theoretical construct. The reason why is straightforward: one cannot generalize from an *n* of one, and the conflict in Kosovo is clearly a “one.”

But this is not to say that *ex post* reconstructions of real world interactions that employ the formal apparatus of game theory are without theoretical merit. For one thing, such reconstructions can rely solely on information that was available to the actors at the time they made their decisions, as Bruce Bueno de Mesquita’s analysis (1998) of the end of the Cold War demonstrates. For another, even if it is true that *some* game model can always be molded to fit the facts of almost any individual event, “closeness of fit” is not the only criterion by which explanatory models are judged. Like a historical description, a game theoretic explanation of an individual case can also be evaluated in terms of the plausibility of its assumptions about the motives of the players and other critical components of the model (i.e., the choices available to the players, the set of outcomes, and so on).<sup>5</sup>

### ■ For What Purpose?

In the literature of international relations, game theory is most commonly used to model *general* political processes such as those associated with crisis bargaining, alliance formation, and war. Here, however, we use it to model a single historical case. What analytic purpose could such a nonstandard application of game theory serve? We submit that there are several.

First, game theory could be used as a strictly *normative* tool to evaluate the efficacy of competing policy prescriptions. For example, Frank Zagare (1984) uses a game theoretic framework to analyze the 1973 crisis in the Middle East, finding that the decision of the Nixon administration to place US strategic forces on alert was not only justifiable, but also more efficacious than the less provocative approach of reassuring the Soviets of a US willingness to compromise.

Game theoretic models could also be employed *descriptively* to explain single cases that are intrinsically interesting or otherwise important (Bates et al. 2000: 700). Game theory is frequently utilized in this way. For example, Zagare’s analysis (1979) of the 1954 Geneva Conference helps to reconcile the well-known but unexplained discrepancy between the public and private policy pronouncements of US decisionmakers both before and dur-

ing the negotiations that ended the Franco-Vietminh war. Zagare (1982) uses the same game theoretic framework to eliminate competing, seemingly plausible, explanations of the Geneva Conference advanced by Ramesh Thakur (1982).<sup>6</sup>

A game theoretic analysis of a single historical case could also be thought of as an inductive step taken to facilitate the development of a general theory. If they are to be useful, formal models, game theoretic or otherwise, cannot be fashioned out of whole cloth. Game theory provides a useful framework for developing comparable case studies, which in turn could serve as a guide in the construction of more refined models or more powerful theories.

### ■ Perfect Deterrence Theory: A Game Theoretic Approach to Deterrence

To summarize briefly, a game theoretic study of a single case can serve a number of theoretical purposes: game theory can, *inter alia*, be used as a descriptive or as a normative device; it can also be employed as a (logical) standard by which competing explanations can be eliminated; and it could be utilized as a tool for theory construction. Beyond these purposes, however, well-developed game theoretic models can be used to shed theoretical light on individual events when an event is identified as an instance of the more general category that is part of the model's empirical domain (Rosenau and Durfee 2000: 3). This is the tack we take in the remainder of this chapter. More specifically, we consider the conflict in Kosovo as an instance of deterrence, and use a set of interrelated game models—called perfect deterrence theory (Zagare and Kilgour 2000)—to gain a deeper understanding of key aspects of the conflict.<sup>7</sup>

Perfect deterrence theory is a theoretical alternative to the standard formulation of deterrence, that is, to classical (or rational) deterrence theory. The two theories have widely divergent empirical expectations and policy implications. These critical differences can be traced to a slightly different axiomatic base. More specifically, classical deterrence theorists like Daniel Ellsberg (1959), Thomas Schelling (1960, 1966), and more recently, Robert Powell (1990) and Barry Nalebuff (1986), assume that conflict or war is always the worst possible outcome of any deterrence encounter. By contrast, in perfect deterrence theory, conflict may or may not be the lowest-ranked outcome of the players.

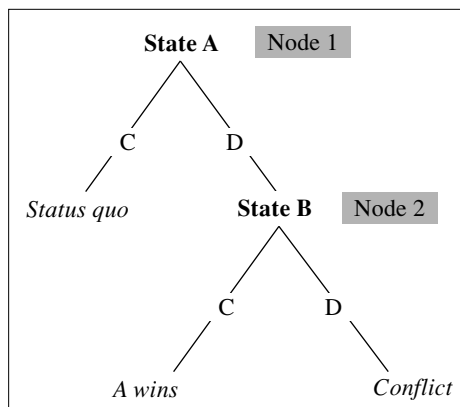
Given that classical deterrence theory was developed at the dawn of the nuclear age, the assumption that all other outcomes are superior to war is certainly understandable. But since this seemingly innocuous assumption leads to untenable logical and empirical conclusions, it is clearly not very useful. To see this, consider for now the *rudimentary asymmetric deter-*

rence game depicted in Figure 4.2.1. In this, the simplest deterrence situation that we can imagine, there are two players, State A and State B, and three outcomes, *status quo*, *A wins*, and *conflict*.

The rudimentary asymmetric deterrence game is a model of an asymmetric or one-sided deterrence situation: State B wishes to deter State A, but not the other way around. Thus, in the extensive-form game depicted in the figure, State A begins play at decision node 1 by deciding whether to *cooperate* (C) and accept the status quo, or to *defect* (D) and demand its alteration. If A chooses C, the game ends and the outcome is the *status quo*. But if State A defects, State B must decide at node 2 whether to concede (C) the issue—in which case the outcome is *A wins*—or deny (D) the demand and precipitate *conflict*. Let us accept, for the moment, two central assumptions of classical deterrence theory: that *conflict* is the worst possible outcome, and that the players are instrumentally rational. What then would instrumentally rational players do when presented with the choices in the rudimentary asymmetric deterrence game?

To answer this question, we examine the game tree of Figure 4.2.1 using a procedure known as *backwards induction*. Specifically, we work backwards up the game tree and determine first what an instrumentally rational State B would do at decision node 2, and then, using this information, specify the rational choice of State A at node 1. At node 2, State B is faced with a choice between conceding (i.e., choosing C), which brings about outcome *A wins*, and denying State A's demand (i.e., choosing D), which brings about *conflict*. But if *conflict* is assumed to be the worst possible outcome, an instrumentally rational State B can only choose to concede, since, by assumption, *A wins* is the more preferred outcome.

**Figure 4.2.1 Rudimentary Asymmetric Deterrence Game**



Given that an instrumentally rational State B will choose to concede at node 2, what should State A do at node 1? State A's choice is either to cooperate, in which case the outcome is the *status quo*, or to defect, in which case the outcome will be *A wins*—because an instrumentally rational State B will choose to concede at node 2. If State A has an incentive to upset the *status quo*—that is, if it needs to be deterred because it prefers *A wins* to the *status quo*—it will rationally choose D. In other words, given two of the core assumptions of classical deterrence theory, deterrence will rationally fail. When joined together, then, the two central assumptions of the standard version of deterrence theory are inconsistent with the possibility of deterrence success.

The inconsistency of these two core assumptions with the stability of the status quo poses a significant empirical problem for classical deterrence theory, namely explaining the stability of the strategic relationship of the superpowers during the Cold War. Recognizing the inconsistency, and attempting to eliminate it, classical deterrence theorists have responded by relaxing the rationality assumption. Schelling, for example, has suggested that a player (such as State B) feigns irrationality to convince another (such as State A) that it would in fact defect at node 2 and execute an irrational threat. Powell, by contrast, holds to the rationality assumption yet allows the threat to be executed, probabilistically, by a disinterested third player, called Nature, that has no stake in the game. Edward Rhodes (1989) abandons the rationality assumption altogether. All of which explains why Christopher Achen argues that “far from leaning too heavily on rational choice postulates, ‘rational deterrence theory’ necessarily assumes that nations are not always self-interestedly rational” (1987: 92).

Unlike classical deterrence theory, perfect deterrence theory assumes that the players are always rational. As well, it drops the assumption that *conflict* is always the worst possible outcome in a deterrence encounter, nuclear or otherwise. In perfect deterrence theory, then, the players may or may not prefer to execute a deterrent threat. It may be surprising to some readers that these two seemingly innocuous axioms lead to empirical expectations and policy prescriptions that are dramatically different than those of classical deterrence theory. But we now show how the implications of perfect deterrence theory provide insights into the Kosovo conflict that simply cannot be derived from the standard formulation of deterrence theory.

### ■ A Game Theoretic Analysis of the Conflict in Kosovo

The origins of the conflict in Kosovo are complex and ancient. They involve both ethnic rivalries and great power political machinations that date back centuries, the details of which need not concern us here. For our

purposes, one need only go back to 1987 and the rise to power of Serbian president Slobodan Milosevic, who skillfully used unrest among the majority of ethnic Albanians in Kosovo for personal political advantage. Soon thereafter, Milosevic terminated Kosovo's autonomy (1989) and abolished its parliament (1990). Of course, Milosevic's repressive policies only made matters worse. In the mid-1990s, moderate Kosovars, such as Ibrahim Rugova, eventually lost influence to more radical elements, the most important of which was the KLA. Increasingly, the secessionist movement in Kosovo turned violent.

A major Serbian crackdown in March 1998 led to open conflict between Serbian military and police forces and the KLA. In short order, over 1,500 Kosovar Albanians were killed and more than 400,000 driven from their homes. The magnitude of the carnage, coupled with a fear that the conflict would spread, led to deeper NATO involvement. NATO's goals were clear: to resolve the conflict by deterring Serbia from continuing the conflict.<sup>8</sup>

Serbia, of course, was not so easily deterred. As John Vasquez notes, "one of the key theoretical questions is why such a small state would be willing to risk war with such a powerful state as the United States" (2002: 107). We address this question presently by bringing to bear the insights offered by perfect deterrence theory. In so doing, we place NATO's conflict with Serbia over Kosovo under what Carl Hempel calls a set of "covering" laws (1965: 345), and then use these laws to explain some of the particulars of that conflict.

To this end, consider first NATO's June 11, 1998, attempt to deter Serbia from further escalating the conflict. On that day, US defense secretary William Cohen directed General Wesley Clark, NATO's SACEUR, to demonstrate NATO's willingness to intervene in the ongoing conflict in Kosovo. Clark responded with a massive show of NATO air power along the Albanian and Macedonian borders. NATO's air exercise, however, failed to end Serbia's five-month counterinsurgency. Why did this specific attempt at deterrence fail? Micah Zenko's explanation is that NATO's threat lacked credibility:

The threat failed because Yugoslav president Slobodan Milosevic recognized cracks in NATO's public display of unity. Before any actual bombing of Serbia could take place, France and Germany insisted that the alliance obtain the imprimatur of a UN Security Council resolution authorizing the employment of "all necessary means" to halt the killing in Kosovo. The United States disagreed, stating that NATO retained the right to act independently of the UN, while Great Britain was on the fence. The allies were aware that efforts to obtain UN support could be vetoed by Russia and China, both judging the violence in Kosovo to be an internal matter for Serbia. Milosevic recognized the operation as nothing more than symbolism, an exercise in public relations, not a rehearsal for war. (2001: 1)

One clear implication of Zenko's explanation is that had NATO's threat been credible, Serbian military operations in Kosovo would have been scaled back and the June 1998 attempt at deterrence would have succeeded. To accept this explanation, however, it must be the case that a credible threat is a sufficient condition for deterrence success. Since perfect deterrence theory reveals that credibility is neither a necessary nor a sufficient condition for deterrence success, Zenko's explanation must be qualified.

To demonstrate why, we must pause and formally define two terms central to any theory of deterrence: *threat credibility* and *threat capability*. We begin with credibility. In the traditional strategic literature, credible threats are oftentimes equated with threats that ought to be believed (e.g., Smoke 1987: 93). Since threats can be believed only when they are rational to carry out (Betts 1987: 12), only rational threats can be credible (Lebow 1981: 15).

The formal definition of credibility in perfect deterrence theory is consistent with the theoretical linkage between threats that are credible and threats that are both believable and rational: credible threats are precisely those that are consistent with rational choice, that is, with threats that the threatener *prefers* to execute.<sup>9</sup> Using the rudimentary asymmetric deterrence game of Figure 4.2.1, it is easy to show why Zenko's explanation of NATO's early attempt at deterrence might appear to be persuasive. We have already shown that the *status quo* is unstable, and deterrence is bound to fail, when State B prefers not to execute its deterrent threat. In terms of our definition, State B's preference for *A wins* over *conflict* renders its threat incredible—assuming, of course, that State A knows State B's preferences.<sup>10</sup>

But now assume that State B's threat is credible, that State B prefers *conflict* to *A wins* and that State A knows this. Consider now State B's node 2 choice in light of this assumption. At node 2, State B can either concede, in which case the outcome is *A wins*, or to deny, in which case the outcome is *conflict*. Given our new assumption about State B's preferences, it is clear that State B, if it is instrumentally rational, would choose to deny State A's demand.

Now consider State A's node 1 choice. If State A chooses to cooperate, the outcome will be the *status quo*. If State A chooses to defect, the outcome will be *conflict*, since State B is now assumed to prefer to execute its deterrent threat. Thus, when State B's threat is credible, the *status quo* is stable, and deterrence will succeed, *as long as State A prefers the status quo to conflict*. If deterrence success depends on both the credibility of State B's threat and State A's preferences, a credible threat cannot be a sufficient condition for deterrence to work. Thus, Zenko's explanation of NATO's failure to deter Serbia from continuing its offensive is inadequate. To be complete, Zenko's explanation would also have to speak minimally to Serbia's preference between the *status quo* and *conflict*.



In fact, in perfect deterrence theory, it is State A's preference between these two outcomes that establishes whether State B's deterrent threat is *capable*. Following Schelling (1966: 7), perfect deterrence theory defines a capable threat as a threat that hurts. Actions that hurt are those that leave a player worse off than if the action were not executed. Operationally, this means that one player's threat is capable only if the other, the threatened player, prefers the *status quo* to the outcome that results when and if the threat is carried out (i.e., *conflict*). In other words, a threat will lack capability whenever the threatened player prefers to act even when a deterrent threat is acted upon. Perfect deterrence theory demonstrates that a capable threat is a necessary condition for deterrence success. This means that deterrence will fail whenever the deterring state's threat lacks capability. It also means that in the absence of a capable threat, the credibility of the deterring state's threat is simply irrelevant. Hence the inadequacy of Zenko's explanation.

At this point the reader might feel as if we are simply picking academic nits. But the distinction between credible and capable threats is both theoretically and empirically significant. To see why, consider a March 1999 meeting between Serbian president Milosevic and US special envoy Richard Holbrooke. At this meeting, which was a last-ditch effort to avoid conflict, Holbrooke asked Milosevic, "Are you absolutely clear in your own mind what will happen when I get up and walk out of this palace?" Holbrooke reported that Milosevic replied, "You're going to bomb us" (McManus 1999: A1). To put this another way, Holbrooke's final threat failed to deter not because it lacked credibility. Indeed, President Milosevic knew that the threat would be carried out. Rather, the threat failed because it lacked capability; Serbia's president preferred to continue fighting in Kosovo regardless of whether or not the threat was carried out. Because the threatened action would not hurt enough, it was insufficient to deter the Serbian offensive.

In the rudimentary asymmetric deterrence game, capability and credibility are jointly necessary and sufficient for deterrence to succeed.<sup>11</sup> Thus it should not be surprising that many of the choices taken by NATO during the conflict were motivated by a desire to establish both (see Chapter 2.3).

### ■ Extended Deterrence and the Conflict in Kosovo

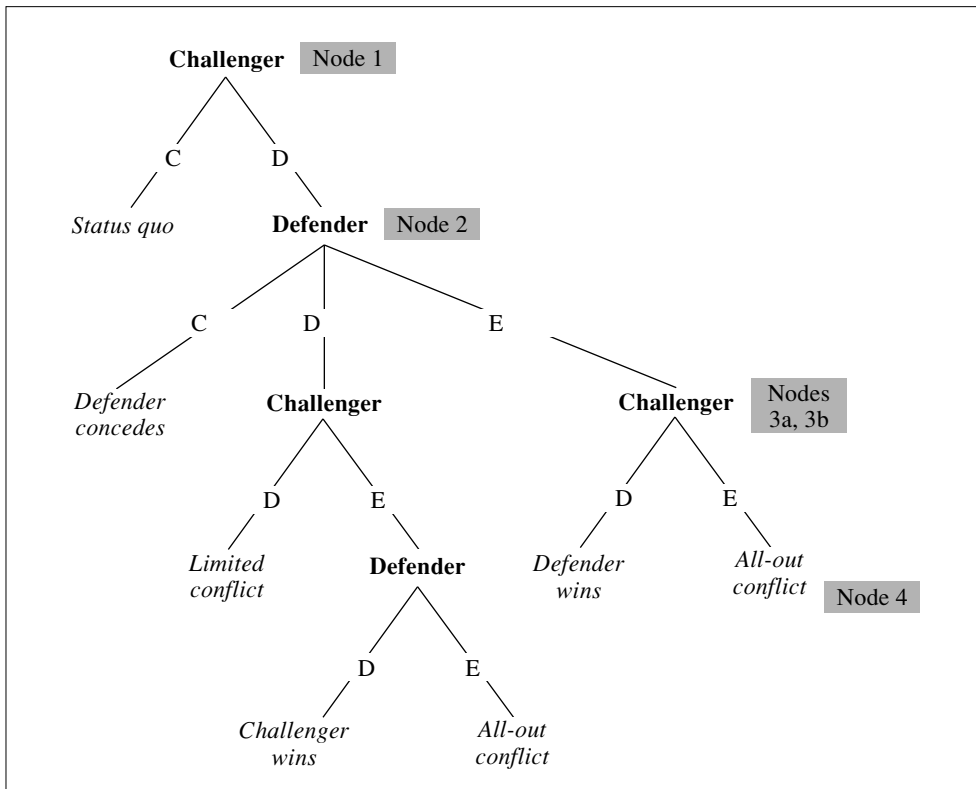
Although the rudimentary asymmetric deterrence game is useful for establishing the importance of credibility and capability (and distinguishing between them), more elaborate game models are necessary to examine the Kosovo conflict in greater detail. By March 1999 the situation in Kosovo was clearly one of extended deterrence. Extended deterrence occurs where one actor, the defender, attempts to deter a challenger from attacking a third

party, the protégé. More specifically in this case, the United States (defender), by way of NATO, was attempting to deter Serbia (challenger) from engaging in ethnic cleansing in Kosovo (protégé).

To model situations such as these, Frank Zagare and D. Marc Kilgour (2000) have developed the *asymmetric escalation game*.<sup>12</sup> As in the rudimentary asymmetric deterrence game, the challenger begins play by choosing whether to cooperate (C) or defect (D). But as Figure 4.2.2 shows, the defender's response options in the asymmetric escalation game are more varied: at decision node 2, the defender can concede (C), it can respond in kind (D), or it can escalate (E). Depending on the defender's choice, the challenger can either escalate first (at node 3a) or counterescalate (at node 3b), or not. If the challenger escalates first, the defender has an opportunity, at node 4, to counterescalate. The outcomes associated with the various choices in the asymmetric escalation game are summarized in Figure 4.2.2.

Given this specification of the alternatives available to each side, the

Figure 4.2.2 Asymmetric Escalation Game



story of the Kosovo conflict can be told in terms of the asymmetric escalation game.<sup>13</sup> In March 1999, despite repeated warnings and threats from the United States and NATO, Serbia continued its campaign in Kosovo; in effect, this meant Serbia was choosing D at node 1. In response, NATO initiated the Operation Allied Force air campaign on March 24, thus also choosing D at node 2. Despite expectations that bombing would lead Serbia to back down quickly, Serbia instead did the opposite, escalating (E) the conflict at node 3a by launching Operation Horseshoe later in March. At this point, NATO decided not to counterescalate; hence ground forces were not deployed.

What led Serbia not only to stand firm in the face of NATO's clear intention to initiate an air campaign, but also to intensify its ethnic cleansing? Some argue that this question cannot be answered, that it is "impossible to know why Milosevic decided to escalate rather than hunker down, let alone capitulate in the manner many expected he would" (Daalder and O'Hanlon 2000b: 93). Through the lens of perfect deterrence theory, however, it is not difficult to explain the Serbian decision to escalate. But one must first recognize that the deterrence attempted by the United States and NATO in March relied on *two* threats: the threat to respond to Serbia's intransigence by an air campaign, and the threat to escalate to a ground war if necessary.

The meeting between Holbrooke and Milosevic on March 22 demonstrated that NATO's threat to bomb Serbia was indeed credible, as discussed above. However, NATO's escalatory threat was just as clearly not credible. And with good reason. As President Bill Clinton stated in his address to the nation on March 24, the United States and NATO did "not intend to put our troops in Kosovo to fight a war" (Clinton 1999c). Similar statements by Vice President Albert Gore, Secretary of Defense William Cohen, British prime minister Tony Blair, and others made it unmistakable that NATO's least-desired outcome was *all-out conflict*.

Given these preferences, one can solve the asymmetric escalation game using backwards induction to determine what outcome should be expected. Given Clinton's obvious signal that NATO would stick with an air campaign *even if Serbia escalated*, that is, that NATO would choose D at node 4, Serbia's node 3a choice is both predictable and easy to explain: after the bombing of Serbian positions, Milosevic decided to escalate because he preferred the outcome that would result if he escalated (i.e., *challenger wins*) to the outcome that would result if he did not escalate (i.e., *limited conflict*). From the Serbian point of view, then, escalation was an instrumentally rational choice.

The United States and NATO faced a more complicated decision at node 2. One thing, however, is incontrovertible: as we have seen, regardless of Serbia's response, a divided NATO was not interested in escalating. But what of the choice of D? In a situation of complete information where each player is fully informed of the other's preferences, NATO would have

recognized that its choice of D at node 2 would lead eventually to *challenger wins*. This may have been less preferred than the *defender concedes* outcome that would have resulted if NATO had chosen to concede (C).<sup>14</sup> However, the United States did not expect Belgrade to escalate. Rather, the expectation in Washington was of a *limited conflict* that would, in short order, be resolved in NATO's favor. As Secretary of State Madeleine Albright stated on national television the night the bombing started, "I don't see this as a long-term operation. I think that this is something . . . that is achievable within a relatively short period of time."<sup>15</sup>

From this it can be inferred that the Clinton administration may have believed that the threat to bomb Belgrade was sufficiently capable and credible to deter Milosevic from fighting on. The belief, while erroneous, was not baseless. Many members of the Clinton administration were convinced that NATO's 1995 air campaign was responsible for Milosevic's capitulation that ended the war in Bosnia-Herzegovina.<sup>16</sup> Another logical possibility is that US decisionmakers did not have complete and accurate information about Serbian preferences, which is simply another way of saying that US intelligence was inadequate. In the end, the conflict did not remain limited as NATO's command had anticipated; rather, Operation Allied Force precipitated an acceleration of Serbia's campaign of ethnic cleansing and led, unexpectedly, to the outcome *challenger wins*.

How then did the situation change, leading Serbia to agree finally to withdraw all its forces from Kosovo and allow entry of an international security force on June 9? The strategic dynamic began to shift after President Clinton made it clear in a press conference on May 18, and a *New York Times* op-ed on May 23, that henceforth NATO would "not take any option off the table."<sup>17</sup> A series of meetings in Washington and in Europe were held in late May and early June to prepare for a possible ground war in Kosovo (Daalder and O'Hanlon 2000b). At about the same time, NATO built up its ground forces in Macedonia and Albania. Unlike the earlier situation, then, the United States and NATO had made sufficiently credible its threat to use ground forces to put a halt to the carnage in Kosovo.<sup>18</sup> Therefore, when it became obvious to Milosevic that Serbia would likely soon be faced with a choice between backing down and a ground war he could not expect to win, he backed down (Erlanger 1999b). In the end, US and NATO forces were able to prevail, but not until after first "losing the war" and then, in the words of Ivo Daalder and Michael O'Hanlon (2000b), "winning ugly."

## ■ Expectations for the Future

While to this point we have been concerned with explaining NATO's failure to deter Serbia in the spring of 1999, perfect deterrence theory can also be used to form expectations for the stability of future relations between

Serbia and NATO. The basic theoretical idea rests on the supposition that some deterrent relationships are inherently more stable than others. If correct, it follows that the specifics of any conflict settlement will condition and shape the future relationship of the combatants and have important consequences for the robustness of the peace (Senese and Quackenbush 2003).

Settlements of militarized interstate disputes fall into two broad categories: imposed and negotiated (Jones, Bremer, and Singer 1996). Winners of disputes that impose settlements are, in general, likely to be quite satisfied with the postsettlement status quo, while losers of those disputes (states that have settlements imposed on them) are likely to be dissatisfied with the new status quo. By contrast, both states in a jointly negotiated settlement are likely to be at once somewhat satisfied with the new status quo (they did, after all, agree to it), but also likely to be somewhat dissatisfied with it (they did not get everything that they wanted).

A state that is dissatisfied with the postsettlement state of affairs will have an incentive to challenge the status quo. To prevent an unwanted challenge, the threatened state will attempt to deter its dissatisfied former opponent from violating the terms of the settlement. Since each side is somewhat dissatisfied following negotiated settlements, mutual deterrence is required to maintain them. However, the maintenance of imposed settlements requires only unilateral (or one-sided) deterrence.

This distinction is important because, other factors aside, unilateral deterrence is easier to establish and maintain than mutual deterrence (Zagare and Kilgour 2000). In consequence, one would expect that imposed settlements to militarized interstate disputes will be more durable than negotiated settlements. Paul Senese and Stephen Quackenbush (2003) confirm this expectation empirically. In an analysis of 2,536 interstate conflict settlements, they find that imposed settlements last significantly longer than negotiated settlements. For example, the expected duration of an imposed settlement is nineteen years. Negotiated settlements, in contrast, are likely to persist for only nine years.

With this theory of recurrent conflict in mind, one can consider the settlement of the conflict in Kosovo.<sup>19</sup> The settlement reached through the Military Technical Agreement of June 9 (which called for the immediate cessation of hostilities and set timelines for the withdrawal of Yugoslav forces from Kosovo) and United Nations Resolution 1244 of June 10 (which welcomed the acceptance by the Federal Republic of Yugoslavia of the principles of a political solution to the Kosovo crisis, including an immediate end to violence and a rapid withdrawal of its military, police, and paramilitary forces) was clearly imposed.<sup>20</sup>

This settlement in large part instituted the provisions of the Rambouillet Accords—which Serbia had previously refused to sign, or

even to seriously negotiate—including the deployment of an international security force (KFOR) to Kosovo to monitor adherence to the settlement. While there is likely to be lingering resentment among some Serbian leaders about these terms, the postconflict peace, and the unilateral deterrence relationship it rests on, are likely to persist, especially if the Federal Republic of Yugoslavia is reintegrated into Europe. This is not to say that a long-lasting peace is guaranteed. But there are both good theoretical reasons and compelling statistical evidence that suggest that the peace will likely endure for some time.<sup>21</sup>

## ■ Conclusion

Game theory is a powerful and versatile methodological tool that lends itself to political analysis in a variety of ways. It is particularly useful for constructing general theories of political interaction. It could also be used as a normative device to evaluate competing policy prescriptions, or as an empirical aid to structure in-depth case studies or even a series of comparable case studies. In this chapter, however, we have taken a less conventional approach. Specifically, we have identified the conflict in Kosovo as an instance of deterrence and then employed a set of interrelated game models, called perfect deterrence theory, to shed theoretical light on a number of aspects of the conflict. In other words, we have taken what we know about interstate conflicts in general in order to achieve a deeper understanding of a singular event.

For example, in perfect deterrence theory, a credible retaliatory threat is neither a necessary nor a sufficient condition for deterrence success. This means, *inter alia*, that the status quo may be contested, and deterrence may rationally fail, even when a defender's threat is in fact credible. Knowing this leads us to question Zenko's explanation (2001) that NATO's attempt to deter a Serbian escalation in June 1998 failed because NATO's threat was not credible. While Zenko may well be right that Serbian president Milosevic did not believe that NATO would use force if he did not deescalate the conflict in Kosovo, it is also likely that, in this instance, Milosevic's beliefs about NATO's intentions were simply beside the point. Indeed, there is considerable evidence that Milosevic intended to continue his Kosovo campaign regardless of whether NATO ordered air strikes. This means that NATO's early attempt at deterrence failed not because its threat was not credible but because it was not capable. In other words, the threat to bomb Serbia did not imply costs that were sufficiently hurtful to alter Serbian behavior.

As a general theory, perfect deterrence theory can also be used to identify the conditions associated with an array of extended deterrence outcomes, including acute crises, limited conflicts, and escalation spirals.

More specifically, crises occur whenever a defender's threat to respond in kind lacks capability or is insufficiently credible to deter aggression. When extended deterrence breaks down, the ensuing conflict is most likely to remain limited when a defender's escalatory threat is both capable and credible. Finally, crises are more likely to escalate when at least one of these two conditions is not satisfied.

The theoretical knowledge of extended deterrence outcomes provided by perfect deterrence theory helps us to provide an answer to what Daalder and O'Hanlon (2000b) claim is an "impossible" question, namely, why Serbia decided to escalate its campaign against the KLA. Using a generic asymmetric escalation game model as a guide, we argue that Serbia's decision to launch Operation Horseshoe was the rational course of action in light of the low credibility of NATO's escalatory threat. Why then did Serbia eventually back down? Clearly, NATO's escalatory threat to deploy ground forces had always been potent (i.e., capable). But when, in early June, it became patently obvious that an invasion was more than just a remote possibility, Serbia suddenly decided that it was in its interest to accede to the demands NATO laid out at Rambouillet. In an uncertain world, brandishing threats that are at once capable *and* credible is the surest path to peace—at least in the short run.

In perfect deterrence theory it is also the case that, other factors aside, deterrence is more likely to be successful, and will be easier to maintain, in unilateral or one-sided relationships than in relationships of mutual deterrence. Knowing this made it possible for us to make a prediction, based on explicit theoretical principles and buttressed by compelling statistical evidence (Senese and Quackenbush 2003), that the settlement imposed on Serbia will take hold.

In sum, our application of perfect deterrence theory to the conflict in Kosovo has enabled us to answer three important questions regarding that conflict. Two of our answers are not only different but are also more complete than answers provided by scholars who have not availed themselves of game theory's powerful analytic tools. We also note that because we have tried to make our discussion accessible to all readers, we have included little in the way of formal proof. Nonetheless, all of the conclusions reached here are grounded in strict logic developed elsewhere. We believe that it is the transparency and consistency of this logic, more than anything else, that makes our conclusions more compelling than the competing explanations we find inadequate.<sup>22</sup>

One criticism of formal theories in political science is that they have not been sufficiently tested and are thus often divorced from reality (Green and Shapiro 1994). Theories of international relations, formal or not, also need to be able to speak to specific real world issues. Recent strides have been made in bridging the divide between formal and empirical studies of

deterrence on a large-sample basis (Quackenbush 2003). In our opinion, this chapter demonstrates that game theoretic models can also be used to gain insight into single cases that are intrinsically interesting, theoretically significant, or both.

## ■ Notes

1. Numerous textbooks provide an overview of the methodology. Among those we recommend are Dixit and Skeath 1999, Fudenberg and Tirole 1991, Gibbons 1992, Morrow 1994, Myerson 1991, Rasmusen 1994, and Zagare 1984.

2. For the distinction between instrumental and procedural rationality, see Zagare 1990. See also Quackenbush 2004.

3. Game theory is frequently criticized for assuming instrumental rationality. But the assumption is often implicit in the analyses of its critics and in what are taken to be antithetical conceptual frameworks. For example, in Chapter 6.2 in this volume, Rosemary Shinko argues that NATO's attempt to frame its intervention in Kosovo as a humanitarian act was *motivated* by a desire to disguise its underlying political character.

4. See also Elster 2000.

5. For a debate on this issue, see Elster 2000 and Bates et al. 2000. See also Ray 1995, chap. 4.

6. For a collection of similar applications, see Bates et al. 1998.

7. Because we are interested in deterrence, we focus on the relationship of Serbia and the NATO alliance. But this is not the only way to view the conflict in Kosovo. One could, for example, also interpret it as an instance of an ethnic conflict or as a civil war. In principle, game theory could be used to analyze those aspects of the conflict. But to do so, one would have to either use or, more likely, develop a general model appropriate to the specific interaction between the Serbians and the Kosovars.

8. Schelling (1966) would call NATO's objective "compellence," that is, a desire to stop an action already under way. For our purposes, the theoretical distinction between compellence and deterrence is not meaningful.

9. Credible threats are those that are consistent with Reinhard Selten's perfectness criterion (1975), that is, with instrumentally rational choices at every decision node of a game tree. Perfect deterrence theory's name comes from its association of credible threats with the perfectness criterion.

10. Technically, we are assuming a game of complete information. We make this simplifying assumption here for the sake of exposition. The general conclusions we reach, however, hold even when this assumption is relaxed. For a demonstration, see Zagare and Kilgour 2000.

11. This statement is not generally true. In some deterrence games, credibility is neither a necessary nor a sufficient condition for deterrence success. Capability, however, is always a necessary condition. For the details, see Zagare and Kilgour 2000.

12. The asymmetric escalation game is a general model. Because it is general, its application to any specific conflict is problematic. In other words, it is meant to apply to a class of events (i.e., situations of extended deterrence) rather than to any single event. We believe that it illustrates a general dynamic of the conflict in Kosovo. But the asymmetric escalation game would have to be modified substantially in order to explore more particular aspects of this or similar conflicts.



13. Our informal analysis of this game is meant to be suggestive. Hence we present only the broad outlines of the model. For a formal analysis that takes payoff parameters into account, see Zagare and Kilgour 2000.

14. However, Vasquez argues that “the United States backed itself into a corner in which the logic of coercive diplomacy made it necessary” to respond to Serbia’s actions rather than concede (2002: 112; see also Judah 2000). Therefore the United States and NATO may have preferred to launch an air campaign rather than concede, regardless of what happened afterward.

15. Madeleine K. Albright, interview, PBS, *Newshour with Jim Lehrer*, March 24, 1999.

16. We are indebted to Jennifer Sterling-Folker for this observation.

17. “Remarks by the President and His Majesty Abdullah II, King of the Hashemite Kingdom of Jordan in Photo Opportunity,” White House, Office of the Press Secretary, May 18, 1999.

18. Andrew Stigler (2002–2003) argues that the threat to introduce ground forces was not credible. But because he treats credibility as a dichotomous variable, Stigler overlooks the possibility that NATO’s threat was *sufficiently* credible to force Serbia to surrender control of Kosovo. In our opinion, a more nuanced analysis is possible when credibility is measured on a continuum, as it is in perfect deterrence theory. For a discussion, see Zagare and Kilgour 2000, chap. 3.

19. Since our argument is supported empirically by an analysis of dyadic *interstate* disputes, we restrict our conclusions to the future relationship between NATO and Serbia. For an analysis of settlements to *intrastate* conflicts, see Hartzell 1999.

20. The text of the military agreement is available online at <http://www.nato.int/kfor/resources/documents/mta.htm>, and the UN resolution text is available at <http://www.un.org/docs/scres/1999/99sc1244.htm>.

21. For a contrary view, see Efrid et al. 2000.

22. For the debate on the importance of logical consistency in theory construction, see Brown et al. 2000.