G

Game Theory and Other Modeling Approaches

Frank C. Zagare and Branislav L. Slantchev

University of Buffalo and University of California – San Diego

Introduction

Game theory, the science of interactive decision making, burst upon the academic scene in 1944 with the publication of von Neumann and Morgenstern's magisterial *Theory of Games and Economic Behavior*. Widely hailed when it was published, this work's impact was felt almost immediately. By the early 1950s, applications and extensions of the original approach began to appear in many of the social sciences, including political science and almost all of its major subfields. International relations was no exception. Indeed, it was in the general area of interstate conflict and its resolution that game theory would make its earliest and most significant contributions.

But that was then; this is now. Much has changed since the first studies were published. Game theoretic models have become increasingly sophisticated and, in consequence, much more powerful and useful. As Walt (1999:5) has written:

> Rational choice models have been an accepted part of the academic study of politics since the 1950s, but their popularity has grown significantly in recent years. Elite academic departments are now expected to include game theorists and other formal modelers in order to be regarded as "up to date," graduate students increasingly view the use of formal rational choice models as a prerequisite for professional advancement, and research employing rational choice methods is becoming more widespread throughout the discipline.

Leaving technicalities aside, what follows is an attempt to trace the evolution of this formal methodology in international relations from its earliest appearance in the scholarly literature to its contemporary application. As one might expect and as will be shown below, these developments closely parallel conceptual refinements and theoretical advances in game theory itself. (See Kuhn 1997 for a compendium of seminal works in game theory.)

Definitions

Before beginning, however, a few introductory remarks about the nomenclature of the methodology are in order. A *game* is any situation in which the choices of two or more actors, called *players*, are interrelated: that is, where the *outcome* does not depend solely on the choice of a single actor. Games are sometimes thought of as lighthearted diversions. And sometimes they are, especially when the stakes are trivial. But it should be clear that a game – as defined above – can be deadly serious. High-stake games are common in international politics.

Games can be categorized along a number of dimensions. One standard division concerns the number of players. *Two player games* are, obviously, games where there are only two decision makers. By contrast, *n-person games* are games played by three or more players. When the players have diametrically opposed interests, a *zero-sum* game is being played; in a *nonzero-sum game*, the players have both competitive and complementary motives.

Games can be further distinguished by the rules that govern play. A *non-cooperative* game is any game in which the players are unable to coordinate their strategy choices. Players may be unable to make coordinated choices for a number of reasons. For instance, they may be unable to communicate, coordination may be precluded by statute, or no agent may exist to enforce an agreed-upon joint strategy. By contrast, a *cooperative game* is any game in which coordinated strategy choices are possible. Since the international system lacks an overarching authority to enforce commitments or agreements, it should come as no surprise to learn that the theory of non-cooperative games holds a particular attraction for theorists of interstate conflict. Bargaining models that rely on cooperative game theory, though important, have been much less influential in the literature of interstate conflict, at least until recently.

Another division concerns a game's information structure. When the payoffs to each player are common knowledge – that is, when they are known to all of the players – the game is said to be a game of *complete information*. In a game of *incomplete information*, at least one player has private knowledge of his or her payoffs.

In general, three conceptual devices have been deployed in the conflict literature to capture the strategic structure of a game: a game tree is used to represent a game in the *extensive form*; a payoff matrix is the basis of the *normal* or *strategic form* of representation; and a mathematical function that assigns a payoff to every player and to every combination of players is known as the *characteristic function form* of representation. The extensive and the strategic forms are typically used in the analysis of two- and, sometimes, three-person games. The characteristic function form is most frequently encountered when an *n*-person game is under consideration.

The First Wave: Zero-Sum Game Theory

Almost all of the early applications of game theory in international relations drew upon the theory of zero-sum games (see, *inter alia*, McDonald and Tukey 1949; McDonald 1950; Williams 1954; Kaplan 1957; Morgenstern 1959; 1961a). Since zero-sum games were the object of von Neumann and Morgenstern's attention, this should not be surprising. But the first generation of applications was also developed during the most intense period of the Cold War. Hence, they also reflected, perhaps unwittingly, the heated political climate in the United States.

Representative of these applications is O.G. Haywood's (1954) analysis of two battles fought during World War II. According to Haywood, the strategies selected by the military commanders in each battle conformed to von Neumann and Morgenstern's famous *minimax solution* for two-person zero-sum games: that is, they were in equilibrium. (Simply stated, this means that all of the battlefield decisions were individually rational; given the opponent's choice, none of the players had an incentive to switch to another strategy. As von Neumann proved in 1928, at least one such equilibrium strategy pair exists in every finite two-person zero-sum game.)

Haywood made much of the conjunction of theory with fact. In his view, his study served not only to "corroborate" the theory, but also to explain the fact. Thus, von Neumann and Morgenstern's approach had both prescriptive and descriptive utility.

The earliest applications of game theory to military affairs, including Haywood's, had little or no impact on the field. Initially, at least, game theory was considered to be a niche methodology in political science, important perhaps in economics or to certain defense intellectuals and strategic planners, but largely irrelevant to the core concerns of international relations specialists. There was, however, one early extension of von Neumann and Morgenstern's framework that was not so easily dismissed: William Riker's (1962) theory of political coalitions.

To develop his theory of coalition formation and disintegration, Riker examined the value of coalitions at the point at which they were just large enough to prevail. After making a number of explicit assumptions about the goal of the players, the value of certain types of coalitions, the possibility of side payments (which are utility transfers between players in a cooperative game), and so on, Riker (1962:32) deduced his well-known size principle: "in *n*-person, zero-sum games, where side payments are permitted, where players are rational, and where they have perfect information, only minimal winning coalitions occur." Riker used the size principle to explain why the three known examples of grand coalitions in the international system fell apart so quickly.

Riker's theory attracted a great deal of attention from international relations theorists with a particular interest in alliances (Dougherty and Pfaltzgraff 1971). Indeed, it seems fair to say that Riker's theoretical contribution greatly accelerated the acceptance of game theory in mainstream political science. It did not, however, immediately bring it about. For this, a distinct second wave in the literature was responsible.

The Second Wave: Nonzero-Sum Game Theory

The theoretical foundations for the second wave of the game theory literature in international relations were, once again, laid by a mathematician. This time it was John Nash, a co-recipient of the 1994 Nobel Prize in economics. (Nash shared the prize with John Harsanyi and Reinhard Selten, whose work will be discussed below.) It was no accident that this prize was awarded on the fiftieth anniversary of the publication of von Neumann and Morgenstern's opus.

Nash's major achievement was to generalize von Neumann and Morgenstern's minimax solution. The result is the now famous Nash equilibrium, the accepted measure of rational behavior in strategic form games. Specifically, a strategy pair is a Nash equilibrium if no player can achieve a better outcome by switching, unilaterally, to another strategy. Nash (1951) showed that at least one such outcome exists in every finite non-cooperative game, nonzero-sum games included.

Nash's conceptual breakthrough brought about a reorientation in formal game theory, but the shift in focus away from purely competitive zero-sum games toward nonzero-sum (or mixed-motive) games was not immediately reflected in applications of the theory in international politics. Nonetheless, the change did occur, eventually, due in no small part to the work of Thomas Schelling (1960; 1966). Indeed, when Schelling's *The Strategy of Conflict* was republished in 1980 by Harvard University Press, Schelling remarked in a new preface that the idea that conflict and common interest were not mutually exclusive, so obvious to him, was among the work's most important contributions. Schelling was awarded the Nobel Prize in economics in 2005. It was well deserved.

But we must be careful here not to give Schelling too much credit. In addition to Schelling's contributions, the shift in the paradigm was encouraged by some technical characteristics of Nash's equilibrium concept that are unique to nonzero-sum games. Two in particular stand out; each posed a distinct and interesting intellectual challenge to those who would apply game theory to international affairs. And, not surprisingly, each inspired a large literature that sought to solve the problem or to minimize its impact.

Even before Nash offered his elegant existence theorem, it was known that an equilibrium pair in a two-person non-cooperative game could be worse for both of the players than a non-equilibrium pair. (Technically, an equilibrium pair could be *non-Pareto optimal*.) That this could be so was more than problematic; it was paradoxical. Since only equilibrium strategy pairs could be considered consistent with individual rational choice, how could two players in a game be better off if they selected strategies that were associated with an outcome that was not in equilibrium? How could two irrational players receive a higher payoff than two rational players?

The game that best illustrates this "paradox of rationality" is game theory's most famous: Prisoners' Dilemma. The name is derived from a story that A.W. Tucker, the chair of Princeton's mathematics department in the 1950s and 1960s, developed to introduce the game's strategic characteristics to his (psychology) students. Tucker's clever story helped to make the game's paradoxical features accessible to those social scientists who were not well versed in the mathematical theory of games. Again, international relations specialists were no exception.

Almost immediately, conflict scholars recognized that the game's strategic dynamic helped to explain why states sometimes engaged in costly arms races that left both no more secure than they would have been had there been no competition (Harsanyi 1965). More generally, the game's structure raised fundamental questions about the possibility of cooperation between and among states in an anarchic system, a question that stands at the center of the dispute between realists and liberal theorists of interstate relations. Realists are quite pessimistic about the possibility of sustained cooperation, while liberals are much more sanguine (Oye 1986). (For the connections between realism and game theory, see Jervis 1988.)

Technical advancements in game theory were part and parcel of this contentious debate. In this regard, two research thrusts deserve special mention. The first is the theory of metagames; the second, the theory of repeated games.

Based on an idea first suggested by von Neumann and Morgenstern (1944:100–6) but more fully developed by Howard (1971), a *metagame* is an imaginary game that is played "in the heads" of the players before they actually make a strategy choice. The metagame itself rests on the supposition that the players are able to anticipate each other's strategy choice, and that each bases its choice on the choice it anticipates the other making. Players in the metagame select metastrategies; stable (i.e. rational) strategy pairs are called *metaequilibria*.

As it turns out, in the Prisoners' Dilemma metagame, conditionally cooperative strategies are associated with one of the game's three metaequilibria, leading Anatol Rapoport (1967), at once a leading contributor to the experimental literature of game theory and a vociferous critic of its application in international affairs (see below), to declare that Howard's theory had provided an escape from the paradox. For Rapoport and others, this meant that interstate cooperation was not only desirable; since it was consistent with rational choice, it was possible as well. Brams (1975:38–9) was not so sure.

Perhaps even more pertinent to the ongoing debate between realists and liberal theorists about the prospects for avoiding mutually hurtful outcomes was the literature on iterated games. Early on, game theorists suspected that individually rational players could cooperate with one another in a Prisoners' Dilemma game that was played repeatedly (Luce and Raiffa 1957:101). But this was not proved. Eventually, the suspicion became part of the folklore of the field, accepted but without formal

		State B: Cooperate	State B: Defect
State A:	Cooperate	*Compromise* (3.3)	*B wins* (2,4)*
	Defect	*A wins* (4,2)*	*Conflict* (1,1)

Key: (x,y) = payoff to State A, payoff to State B
4 = best; 3 = next-best; 2 = next-worst; 1 = worst
* = Pure strategy Nash equilibrium

Figure 1 Chicken

demonstration. But during the second wave Taylor (1976) was able to show, albeit for a restricted set of strategies, that mutual cooperation was consistent with rationality in a repeated Prisoners' Dilemma game, provided that each player's discounting of future payoffs was "sufficiently" low. And Axelrod (1981; 1984), taking an evolutionary approach, found that repetition made it possible for mutual cooperation to emerge in a world of non-cooperating egotists, provided that individuals in a cluster of cooperating players interacted with one another more than they did with non-cooperating members of the general population. Drawing different conclusions from these results, neoliberals and neorealists continued to debate whether the glass was half-full or half-empty (Keohane 1984; Baldwin 1993).

In addition to Prisoner's Dilemma, the nonzero-sum game of "Chicken" also figured prominently in the literature of interstate conflict during the second wave (Snyder 1971; Freedman 1989: ch. 12). There are three Nash equilibria in Chicken. One is in mixed strategies; but two are pure strategy Nash equilibria. (For a discussion of the distinction between pure and mixed strategies, see Zagare 1984.) Notice from Figure 1 that the pure strategy Nash equilibria in Chicken are not *equivalent*: that is, that they yield different payoffs to the players. Also notice that the strategies associated with the equilibria are not *interchangeable*: that is, they do not always lead to the same outcome or payoff. Multiple nonequivalent and/or non-interchangeable pure strategy equilibria are unique to nonzero-sum games; Nash equilibria in zero-sum games are always equivalent and interchangeable.

For game theorists, the properties of multiple equilibria in nonzero-sum games simply confound analysis. But for conflict theorists these properties raised yet another impediment to interstate cooperation: policy coordination (Stein 1983). In other words, in an anarchic world, it may be difficult for states to synchronize their strategy choices, even when they have identical interests. Interdependence and a commonality of purpose do not automatically lead, rationally, to mutual cooperation.

While game theorists proposed formal solutions to the "coordination problem," Schelling (1960) suggested that focal points, or prominent features of the strategic landscape, could facilitate tacit agreements between states. Two states, for example, might both hold back from crossing a physical, political, or psychological boundary because, once crossed, there was no other obvious point of agreement. Schelling's creative solution, however, did not gain general acceptance. Focal points, many thought, were in the eye of the beholder and, hence, extra-game-theoretic (Riker and Ordeshook 1973:226–7).

Chicken is obviously a dangerous game to play. To "win" this game (by choosing to Defect) a player must risk *Conflict* – its worst outcome. But cooperation is also a perilous choice; a player seeking compromise risks being exploited by its opponent – its second-worst outcome. Neither choice could be ruled in – or out – by the dictates of rationality.

At a time when game theorists were asking how this tricky game should be played, and strategic planners were wondering how best to "manage" a crisis, some conflict theorists – Young (1975) calls them manipulative bargaining theorists – concluded that these questions were identical. Heavily influenced by an elementary decision-theoretic model developed by Ellsberg (1959), these theorists developed a counterintuitive tactical approach to foreign policy. Schelling's (1960; 1966) work is, once again, seminal here, but important contributions were also made by Kahn (1960; 1962; 1965), Snyder (1961; see also Snyder and Diesing 1977: ch. 6), Jervis (1972), and others. The stratagems they offered – seizing the initiative, feigning irrationality, linking seeming extraneous issues to one another, forfeiting control, and so on – were both novel and counterintuitive. Jervis (1979:292) would later admit that they were "contrary to commonsense." Some game theorists were harshly critical of what they considered to be an abuse of the methodology (Morgenstern 1961b; Rapoport 1964; 1968). Recently, the empirical foundations of manipulative bargaining theory have also been called into question (Huth 1999; Danilovic 2001; 2002).

Crisis bargaining theory was not the only part of the literature of interstate conflict to be heavily influenced by the Chicken analogy. The theory of mutual deterrence was also deduced from the structural characteristics of this strategically bedeviling game. Though each literature is distinct, the lines between them were oftentimes blurred. For their part, crisis bargaining theorists sought to prescribe "winning" stratagems during a conflict. By contrast, deterrence theorists focused on avoiding confrontations altogether.

It is important to note that the two bodies of literature rested on entirely different suppositions. Manipulative bargaining theorists presumed that, during a crisis, the threat of nuclear war could be used for political advantage. Deterrence theorists took as their point of departure the absurdity of nuclear war. Since these two ideas are "fundamentally inconsistent" with one another, it should not be surprising to learn that the strategic literature of the period was characterized by what Trachtenberg (1991:32) calls a "pervasive" and "persistent" tension between them. What should be surprising to learn, however, is that the two contradictory theories were developed by the same set of defense intellectuals, who drew on the same analogy, who worked in the same paradigm, and who were obviously trying to have it both ways.

The idea that the superpower relationship of the Cold War era could be modeled by Chicken came under intense criticism. Some thought the stark structure of a 2×2 matrix could not capture the subtlety of such a complex strategic situation. Others thought that the serious business of interstate conflict should not be considered a mere "game." But the theory of mutual deterrence, as initially developed, had a more fundamental (technical) problem, a problem that Chicken very nicely illustrates: in light of the fact that the compromise outcome in Chicken is not a pure strategy Nash equilibrium (that is, is inconsistent with rationality), how can the absence of a superpower war be explained?

Schelling's answer/prescription was the "threat-that-leaves-something-to-chance." More specifically, the stability of the Cold War status quo could be explained by each side's fear that even a minor challenge to the existing order might set off an escalation spiral that neither side could control. In other words, cooler heads prevailed simply because decision makers in both Washington and Moscow were afraid of an "inadvertent" or "accidental" war. The soundness of this extremely influential strategic argument (Trachtenberg 1990/1:120) was hotly debated during the literature's third wave.

One persistent tension in these early applications of game theory to international politics was the mismatch between the verbal descriptions that analysts offered and what the models actually represented. If there was one feature of static (i.e. strategic form) games that tended to limit their usefulness, it was that they were, well, static. They appeared ill-suited to address inherently dynamic situations like international crises where moves and countermoves determine the eventual outcome of the interaction (Wagner 1983).

The concern with dynamics was clearly evident in the approach of Fraser and Hipel (1979), an extension of Howard's (1971) analysis of options technique and a subtle attempt to redefine Nash's notion of an equilibrium strategy pair/outcome. Their methodology involved a listing of all relevant player strategies and a stability analysis of all feasible outcomes. Stable outcomes (i.e. equilibria) were those that were individually rational for every player. But in defining rationality, Fraser and Hipel assumed that the players not only took into account the consequences of an immediate strategy switch, but also considered the likely response of the other player or players. Thus their definition of rationality was more demanding than Nash's.

Brams' (1994) theory of moves was another attempt to account for dynamics. In developing this framework, Brams started with a static 2×2 strategic form game. From this he derived an extensive form game in which the players consider moves, countermoves, counter-countermoves, and so on. Stable or rational outcomes in this context are referred to as *nonmyopic equilibria* (Brams and Wittman 1981). They are stable because the players, looking ahead and evaluating the long-term consequences of a strategy switch, realize that there is nothing to be gained by doing so. Unlike Fraser and Hipel's definition of an equilibrium outcome, the concept of a nonmyopic equilibrium places no arbitrary limitation on the number of moves and countermoves the players can make or consider. Zagare's (1987) first attempt to develop a general theory of deterrence relied on Brams' provocative attempt to recast classical game theory.

While some scholars were trying to develop the game-theoretic tools to make them suitable for the analysis of dynamic situations, others attempted to represent such situations with action–reaction models of behavior. In this approach, actions are responses to stimuli in the environment: what actor A does at time t is largely dependent on what actor B did at time $t - 1$, which was itself largely dependent on what A did at $t - 2$, and so on. The most famous application of this method was Richardson's (1960) model of arms races in which a nation's increase in arms is a function of its current strength, the costs of arming, the level of hostility toward its opponent, the opponent's military strength, and the nation's willingness to arm because of that strength. Richardson formulated a pair of differential equations that captured this interaction, estimated a statistical model using defense spending data from the nineteenth and twentieth centuries, and found significant evidence of reciprocity.

The methodology seemed to offer an ability to deal with dynamic processes and, just as importantly, provided an opportunity to integrate theory and data analysis in a more rigorous fashion than the alternative approaches. Richardson's pioneering work ushered in nearly two decades of intensive research employing the action–reaction methodology, most of it empirical, and much of it devoted to arms races (Brito 1972; Majeski and Jones 1981). Somewhat surprisingly, Richardson's original finding was among the first victims, with studies finding little evidence that a nation's arms buildup was primarily a response to an increase in armaments by a rival (Hollist 1977; Ostrom and Marra 1986), although others fault the methodology for failing to uncover reactivity (Ward 1984).

More recently, Kadera (2001) constructed a differential equations model that integrated balance of power and power transition (preponderance of power) theories. She assumes that each is right at some point of time, which is in contrast to Powell's (1996) argument that the theories are missing an essential "intervening" variable in the status quo distribution of benefits. Starting from Doran and Parsons' (1980) claim that power evolves over time according to an S-shaped curve similar to population dynamics, Kadera developed a set of equations that captured this dynamic and allowed for the possibility of conflict. She studied unsuccessful transitions, temporary transitions in which the dominant state manages to recover its position, and successful transitions in which the rising challenger overtakes the declining state permanently.

Despite the early promise of this approach, its mechanistic rigidity has been its primary shortcoming (McGinnis 1991). Although the actions in an action–reaction model are interdependent, they are not the result of choices because actors do not evaluate the consequences of their behavior or attempt to optimize. Thus, the action–reaction approach introduced dynamics but at the expense of strategic choice. As Richardson (1960:12) himself put it, this approach captures "what people would do if they did not stop to think." What analysts needed was a tool that would incorporate dynamics while retaining the notion of rational choice: that is, a development of game theory.

The Third Wave: Dynamics and Equilibrium Refinements

During the third wave, formal modelers began to think outside the (2 × 2) box. From roughly the early to mid-1980s to the mid-1990s, there was a distinct move away from static strategic form games toward dynamic games depicted in extensive form. The assumption of complete information also fell by the wayside; games of incomplete information became the norm. Technical refinements of Nash's equilibrium concept both encouraged and facilitated these important developments.

At about the same time as these disciplinary trends were emerging, there was a not entirely unrelated revival of interest in rational choice modeling in all areas of political science, including security studies and international political economy (Zagare 1990b). The literature of the second wave was partially responsible for the dramatic shift in the political science literature of the 1990s. But in international relations, the publication of Bueno de Mesquita's (1981) *The War Trap* all but sealed the deal. Eventually, attempts to refine, extend, and apply his expected utility theory of interstate conflict initiation would converge with the efforts of those who were dissatisfied with static game-theoretic models (e.g. Bueno de Mesquita and Lalman 1992).

Despite their inventiveness, neither Fraser and Hipel's (1979) nor Brams' (1994) attempts to introduce dynamics into game theory became standard. The refinement of Nash's equilibrium concept that did gain wide acceptance among game theorists is called *subgame perfect equilibrium*. Due to the ground-breaking efforts of Reinhard Selten (1975), subgame perfection is now the accepted measure of rational play in dynamic (extensive form) games of complete information.

The idea behind this important concept is both simple and intuitive. A subgame perfect equilibrium requires that players make rational choices at every node of a game tree: that is, at each and every opportunity they have to make a choice in a game. This requirement, absent in the definition of a Nash equilibrium, ensures that only outcomes supported by credible (i.e. rational) threats are considered stable. All outcomes that satisfy the perfectness criterion are Nash equilibria, but not the other way around. Thus, Selten's redefinition eliminates from consideration precisely those Nash equilibria that are inconsistent with individually rational choice and enhances, considerably, our understanding of the concept of instrumental rationality (Zagare 1990a).

The natural extension of the perfectness criterion to dynamic games of incomplete information is called *perfect Bayesian equilibrium*, an important development which can be traced to the pioneering work of John Harsanyi (1967; 1968a; 1968b). Briefly, a perfect Bayesian equilibrium consists of a plan of action (i.e. a strategy) for each player, plus a set of beliefs about (i.e. subjective probabilities over) the other player's type (one for each player), such that each player (1) always acts to maximize its expected utility given its beliefs, and (2) always updates those beliefs rationally (i.e. according to Bayes's rule) given the actions it observes during the play of the game.

Subgame perfect and perfect Bayesian equilibria entered the conflict literature during the third wave, slowly at first, but eventually these concepts and their associated game forms almost completely eliminated applications of (mostly 2×2) strategic form games to international security affairs. In this regard, Powell's (1987; 1988; 1990) work on deterrence was at the forefront of this dramatic conceptual transformation of the field.

Building on the earlier, second wave literature, Powell's models postulate a deterrence relationship in which (nuclear) threats were inherently incredible, as in Chicken. He then asked if the absence of a superpower conflict during the Cold War could be reconciled with individual rationality.

To address this question Powell developed a sequential game model in which the players are not in full control of the outcome. Specifically, an escalatory move by either player unleashes an autonomous risk that a conflict will, unintentionally, spiral to the highest level. Thus, his model provides a formalization of Schelling's (1960) "threat-that-leaves-something-to-chance." In so doing, it captures well the classical view that nuclear crises are "competitions in risk taking." Significantly, Powell showed that conditions existed under which neither player would contest the status quo, thereby resolving in the minds of many the paradox of mutual deterrence.

Powell (1987:725), however, was somewhat more circumspect in his conclusions. He notes that he agreed with the point "that requiring the states' strategies to be sequentially rational and then relying on Nature to impose the irrational sanction does not really solve the credibility problem." Powell went on to note, rightly in our opinion, that "it is important to realize that this is not so much a criticism of the model as it is a fundamental criticism of the way that the strategy-that-leaves-something-to-chance has attempted to overcome the credibility problem. The model only exposes this weakness."

To reconcile rationality with deterrence more satisfactorily, Zagare and Kilgour (2000) developed perfect deterrence theory. Their general theory, which takes part of its name from Selten's equilibrium criterion, explores deterrence relationships in the context of a number of interrelated incomplete information game models. It is important to point out that perfect deterrence theory's axiomatic base differs from the standard formulation's. Previous models of deterrence, including Schelling's (1960), Powell's (1990), and Nalebuff's (1991) start with fixed preference assumptions that mirror the preferences of the players in Chicken. Since conflict is a mutually worst outcome in Chicken, deterrent threats are always irrational to carry out in these classical models, which explains why Powell relies on an autonomous force (player) to execute them probabilistically. By contrast, in perfect deterrence theory, the credibility of threats is not fixed. Additionally, only credible threats can be executed, and only by the players.

Zagare and Kilgour claim that their re-specification of deterrence theory eliminates the logical inconsistencies that undermine the classical formulation. More specifically, they show that, in an uncertain world, mutual deterrence may be consistent with instrumentally rational choices even when both players have less than fully credible retaliatory threats. Zagare and Kilgour argue that, unlike classical deterrence theory, perfect deterrence theory makes consistent use of the rationality postulate; that it is *prima facie* in accord with the empirical record; and that its commonsense policy prescriptions are grounded in strict logic (Zagare 2004).

The Latest Wave

In the fourth and final wave, which can be dated, roughly, from around the middle of the 1990s, extensive form games of incomplete information appeared regularly in the strategic literature. Subgame perfect and perfect Bayesian equilibria were no longer considered novelties. Game-theoretic applications proliferated and made important contributions in a number of substantive areas (O'Neill 1994a; 1994b; Snidal 2002 provide detailed reviews). Walt (1999), however, expressed the procedurally irrational fear that models that were based on the concept of instrumental rationality were in the process of taking over the security field. His intolerant call for intellectual tolerance left many choice theorists perplexed, to say the least (Bueno de Mesquita and Morrow 1999; Martin 1999; Niou and Ordeshook 1999; Powell 1999; Zagare, 1999). As should be clear from this sometimes contentious exchange of views, the fourth wave is a period in which game theory was no longer considered a niche methodology. Indeed, it emerged as a mainstream theoretical tool.

This period has seen two major parallel activities: elaboration of existing ideas and innovation, both in the questions being asked and in the formal tools brought to bear on the analysis. Initially, formal models tackled existing controversies and attempted to bring discipline to the myriad of loose, competing, and contradictory arguments floating in the literature. What some critics saw as merely "pouring old wine in new bottles," has been an activity that has enriched our knowledge and deepened our understanding of the phenomena under investigation. The value added of this work is significant: it exposes assumptions necessary to sustain some arguments, shows how other conclusions depend on unstated assumptions, unifies previously incompatible arguments within the same framework (so informal conclusions dependent on contradictory assumptions are now obtained in the same framework), and exposes the instability of some conclusions by noting their sensitivity to the precise specification of assumptions. Although perhaps not original in the sense that the questions it tackled were new, this work was original in the sense that it made previously unknown contributions to knowledge.

Two controversies that received formal attention are the relative gains debate and the question about the relationship between the distribution of power and the probability of war. Scholars in the neoliberal tradition argued that international institutions can help overcome the problems of international anarchy and promote cooperation (Stein 1983; Keohane 1984). Those working in the realist tradition disagreed: because states can always resort to arms, they will be greatly concerned about how the gains from cooperation are distributed among them. In other words, the concern about relative, rather than absolute, gains will diminish the incentives to cooperate (Mearsheimer 1990; Grieco 1988). Formal theorists showed that the debate about state preferences – whether states maximize absolute or relative gains – is misplaced. Snidal (1991) demonstrated that even if one assumes that states maximize relative gains, they will cooperate unless one also assumes there are only two of them and that the distribution of absolute gains creates a Prisoner's Dilemma. Powell (1991) started from the opposite assumption – that states maximize absolute gains – and found that strategic interaction produces behavior that makes them appear as if they care about relative gains.

Scholars also generally disagreed about how the distribution of power is supposed to affect the likelihood of war. The two dominant schools of thought are the *balance of power*, which holds that the more evenly power is distributed, the less likely war is (Morgenthau 1948; Mearsheimer 1990; Wright 1965), and the *preponderance of power*, which holds the exact opposite – namely, that the more asymmetric the distribution of power is, the less likely war is (Organski 1968; Organiski and Kugler 1980; Blainey

1988). Although some formal results suggest that there may be no direct relationship between pre-crisis indicators of power and the probability that escalation ends in fighting (Wittman 1979; Fearon 1994b), the most influential re-analysis of this issue is provided by Powell (1996). He looked at a crisis as a problem of bargaining between two actors who are asymmetrically informed about each other's expected payoffs from war and then explored when this negotiation is likely to break down with one of the actors attacking the other. Powell (1996:241) found that the distribution of power is only related to the probability of war insofar as it is highly discrepant from the existing distribution of benefits: "the probability of war is a function of the disparity between the status quo and the distribution of power." Thus, war may be very likely when the actors are nearly evenly matched in power but one of them benefits disproportionately from the status quo (contradicting balance of power), but it can also be very likely when there is a serious asymmetry in power but the benefits are approximately evenly distributed (contradicting preponderance of power). These theoretical hypotheses, which are yet to be fully tested, provide a ready explanation for the inability of large-*N* empirical studies to find a systematic relationship between the distribution of power and war.

This brings us to a separate consideration of the bargaining model of war, which is perhaps the best example of cumulative knowledge in an entirely new area. The literature on this is now quite extensive, so only a cursory glance is possible. Although the number of works devoted to the "causes of war" is enormous, it was not until Morrow's (1989) and Fearon's (1995) seminal articles that the puzzle could be fruitfully framed in rationalist terms. This approach begins with a simple observation: namely, that war is very costly, often exceedingly so. But if Clausewitz is right and war is just another means of resolving political disputes, then using force to settle them seems an incredibly risky and wasteful way to go. In particular, as Fearon shows with a very simple model, one can always find agreements that would leave *both* actors better off in expectation than waging war. The puzzle, then, is why these actors are unable to find such an agreement and avoid fighting. In this view, to explain war one must explain bargaining failure. Of course, Schelling (1966) himself talked about war as a bargaining process, and Kecskemeti (1958) made an analogous argument in his book. But it was not until the formal elucidation that the puzzle came starkly into view.

The impact was immediate and widely felt. Game theorists interested in conflict quickly proceeded to investigate the robustness of the various explanations that Fearon (1995) offered for bargaining breakdown. The simple model was attacked as being unrealistic, particularly because it did not allow for more extensive bargaining – negotiations were limited to one side delivering an ultimatum to the other. Powell (1996), however, demonstrated that even if one allows the two sides to alternate making offers without any particular time horizon to their interaction, one of the major explanations emerged intact. This became known as the "risk–return trade-off," and it states that when one or both actors are uncertain about how much their opponent expects to gain from war, then it may be impossible to find a mutually acceptable deal at the bargaining table. The fundamental reason is that an actor would not want to give up a lot more than the other would accept in order to forgo war, but each actor has an incentive to ask for as much as possible. The optimal course of action then is to balance the risk of having one's demand rejected (and having to fight a war) with the gain of obtaining better terms if it is accepted. Recently, however, the robustness of this conclusion was challenged by Leventoglu and Tarar (2008), who showed that if the structure of the model is modified slightly in an intuitive way, the uniqueness of the risk–return trade-off equilibrium disappears. Instead, they insisted that one should look at alternative, less risky but also costly, ways of screening out one's opponents at the negotiation table.

The notion of conveying information during bargaining is, of course, not new. However, crisis bargaining – that is, negotiations in the shadow of power – are made much more complex by the ever-present option of quitting the bargaining and going to war. Banks (1990) showed that with asymmetric information, crisis bargaining will almost invariably involve the risk of breakdown. In fact, he proved that in a wide class of models, an actor who expects to gain a lot from war can also expect to get a better deal if negotiations succeed, but at the cost of running a higher risk of war. This equilibrium result justified the earlier assumption (e.g. Bueno de Mesquita 1981; Bueno de Mesquita and Lalman, 1986) that stronger actors would run higher risks because it *derived* it from incentive-compatibility primitives. But research did not stop there: building on Jervis's earlier non-formal work on signaling, Fearon (1997) offered two ways through which actors may reveal private information during a crisis. One entails incurring costs irrespective of the outcome (sinking costs) and the other entails incurring costs if the actor fails to follow through on his threats (tying hands). The second mechanism received a lot of attention. Fearon (1994a) showed how leaders who can create *audience costs* for backing down after making threats can credibly commit themselves to war and thereby obtain better negotiated deals. Schultz (1998) elaborated on this argument by showing that the presence of a political opposition along with institutions that allow it to express its position on the dispute can also enhance the credibility of the leader's threats. Some scholars challenged the microfoundations of the audience cost assumption because it is not immediately apparent what sort of rational process would cause the public to sanction their leaders for bluffing during a crisis (Smith 1998; Schultz 2001; Slantchev 2006).

The consensus in the crisis bargaining literature seemed to be that for a signal to be informative it must be costly, and that the costliness must come from the increased risk of war. Slantchev (2005), however, argued that military threats are not pure instances of either cost sinking or hands tying. Instead, they share characteristics with both because although they must be paid for regardless of how the crisis turns out, they also increase one's preparedness of war and therefore can serve as a commitment-creation device. He showed that stronger actors can still obtain better negotiated deals but do not actually have to run a higher risk of war to do so. Somewhat surprisingly, the military instrument can involve *lower risks* than diplomatic exchanges.

The idea that diplomatic exchanges can be fruitful even when they are confined to mere words is quite controversial, but it does have its own proponents. Sartori (2005), for instance, showed how threats can be credible if being caught lying would incur long-term reputational losses because other actors ignore a "liar's" communications for a while. Kurizaki (2007) revealed a surprising rationale for private communications. While the conventional wisdom holds that for a threat to be credible it would have to be made in public, Kurizaki found that even private threats can work because they allow the opponent to back down in private without engaging his reputation and honor.

For all this emphasis on finding ways to reveal credibly private information, recent work reminds us that one must not neglect instances where strong actors may have incentives to pretend to be weak. Slantchev (2007) showed that an actor may prefer to feign weakness during a crisis if doing so would improve his chances in the war that would follow should bargaining break down.

Unlike most other work on the causes of war, the formal literature tackled explicitly the Clausewitzian notion of bargaining through fighting. The modern models now regularly incorporate war fighting to investigate how actors can terminate war and how peace agreements can hold up afterwards (Wagner 2000; Filson and Werner 2002; Senese and Quackenbush 2003; Slantchev 2003; Powell 2004; Smith and Stam 2004). Using this type of model, Leventoglu and Slantchev (2007) offered one possible solution to the puzzle of war in the absence of uncertainty: they demonstrated how the destruction of value during war can improve the chances of forging an enduring peace settlement.

In addition to "unpacking" the process of war itself, scholars also relaxed another assumption common to formal and non-formal models alike – that of the unitary actor. The tradition of treating states as individuals is long but nonetheless vexing. Formal theorists explored several venues. First, they looked at how the personal incentives that leaders face may cause them to undertake actions contrary to the interests of the public. Second, they analyzed how domestic institutions shape political decisions and affect the probability of war.

The first line of research is based on the principal–agent models common to economics (Morrow 1991). In these, the principal (the public) selects an agent (the leader) to act on its behalf. The agent is assumed to be more knowledgeable about the area where he is supposed to act, and to have access to better information. The problem is that his preferences, and sometimes even his actions, cannot be observed directly by the public, which has to judge his performance on the basis of observable outcomes. This introduces a discrepancy between what the principal wants the agent to do and what is in the agent's own interest. The goal is to provide the agent with appropriate incentives to make him behave how the principal wants (Downs and Rocke 1994). A good example of the benefits offered by formalization is provided by the large literature on the so-called "Schelling conjecture," the informal idea that domestic constraints (e.g. a ratification procedure) can provide an actor with leverage at the international bargaining table (Schelling 1960). Putnam's (1988) article revived interest in that notion, but it was the formalization that explored the conditions under which it can be expected to hold (Iida 1993; Mo 1995; Milner 1997). These studies have shown that the intuitively plausible logic of the conjecture does not hold in many circumstances. For instance, under incomplete information domestic constraints could make both negotiators worse off: that is, the exact opposite of the conjecture (Tarar 2001).

The selectorate theory developed by Bueno de Mesquita et al. (2003) begins by assuming that leaders can remain in power only by retaining the support of the minimal group that is sufficient to keep them in office, the winning coalition. The winning coalition in a democracy is close to one half of the population, whereas in an autocracy it is limited to a much smaller group of people. The leader must provide members of this coalition with a mix of private and public goods that will make them willing to support him in office. Because the size of the coalition differs across regime types, the theory can make predictions about the mix of benefits that leaders will provide in different institutional settings (the larger the winning coalition, the more public benefits are provided), among other things. The theory has dealt with the outbreak of war, war aims, and war termination as well.

Formal studies also broke new ground in studies of missile defense systems and protection against terrorist attacks. Powell (2003) continued his work on deterrence theory by looking at the effect of nuclear proliferation on the ability of the United States to pursue its foreign policy interests, and how a missile defense can offset the adverse effects of proliferation. He argued that the more "roguish" the opponent is, the worse the US will be in the nuclear brinkmanship interaction, and that a missile defense system may actually increase the risk of a nuclear attack upon the United States because it may increase US willingness to press the crisis. Quackenbush (2006) disputed this conclusion and argued that instead of classical deterrence theory, which he criticized for its reliance on Schelling's (1966) "threat-that-leaves-something-to-chance," one should apply Zagare and Kilgour's (2000) perfect deterrence theory instead. He found that, contrary to Powell's (2003) pessimistic conclusion, a national missile defense system should generally enhance the stability of deterrence, although it may increase the dissatisfaction of potential challengers.

In his work on defense against terrorist attacks, Powell (2007b) noted that when the defender is privately informed about the vulnerability of its sites, he must balance the better protection afforded by higher investment in defense of particularly vulnerable sites and the higher appeal such sites have to potential attackers, who can infer their vulnerability from the increased spending. In a related article, Powell (2007a) studied how the defender might allocate limited resources to defend numerous sites when it is uncertain which ones might be targeted by terrorists. The analysis provides an algorithm for optimal allocations which may involve site-specific defenses and general defenses such as border protection or investment in counterterrorist and intelligence operations.

A recent addition to the analyst's formal modeling toolbox are the agent-based models that use computer simulations to investigate the aggregate behavior of complex systems comprising multiple agents. Cederman's (1994) geopolitical model is a good example of this approach. Here, individual actors are endowed with resources and behavioral characteristics, and then after setting the initial conditions for the entire system, these actors are allowed to interact over some period of time. The focus is on long-term dynamics: the patterns (e.g. territorial distribution) that emerge, and their stability. The advantage of this approach over game-theoretic models is that it permits the analysis of a collection of actors over a period of time, something that is either exceedingly complicated to do in a standard game-theoretic model or that yields indeterminate results (e.g. Folk Theorems where almost every pattern of behavior can be supported in equilibrium with sufficiently complex strategies). The disadvantage is that agents usually have limited look-ahead, if any, and as such do not behave strategically. The behavioral assumptions lack microfoundations, and the system is very sensitive to initial conditions. The most exciting venue for future research is combining game-theoretical interactions with computer simulations to study situations with many agents that do not permit closed form solutions.

Finally, game-theoretic work itself began to bridge the gap between "hard" rational choice and "softer" traditions. O'Neill's (1999) work on understanding the symbolism of words and actions of actors is exemplary. He used modified game-theoretic models to deal with concepts such as honor, prestige, face-saving, and moral authority to explain a wide range of phenomena (e.g. arms control agreements, responses to crisis tensions) without resorting to assumptions about emotions, rhetoric, or psychological pathologies. Instead of looking at how messages can be made credible – the focus of traditional models – he asked how one can understand the meaning of these messages. This is very new, and it remains to be seen just how influential it will become.

To conclude, it is perhaps best to temper the upbeat tone of this essay with a discussion about some of the commonly acknowledged limitations of game theory. Although a full critique is well beyond the scope of this essay, it is necessary to acknowledge that game theory is more than a neutral tool for analysis because it carries a host of assumptions, some of which may be more troublesome than others. For instance, the very notion of a Nash equilibrium requires a common conjecture about strategies or else it becomes impossible to explain why actors would choose the equilibrium strategies (Aumann and Brandenburger 1995). This demanding assumption has led to the definition of weaker solution concepts like rationalizable strategies (Bernheim 1984; Pearce 1984), and correlated equilibria (Aumann 1987), and, more recently, to a movement away from the notion of equilibrium models of learning where behavior is based on past experience and limited forward looking (Fudenberg and Levine 1998).

A hotly debated and unresolved problem also arises when the very intuitive notion of backward induction (and the closely related concept, subgame perfection) is used in some games like the famous Centipede game where the solution involves stopping the game immediately even though continuing for a while would be better for both players. The subgame perfect solution strikes many as not something that rational players would actually do, an intriguing case of hyperrationality leading to irrationality (Binmore 1987), and has stimulated research into the role of counterfactuals in strategic reasoning (Bichieri 1988).

Another problem is that equilibrium play depends on off-the-path conjectures about behavior, and analysts have had difficulty deciding what an actor must believe if an out-of-equilibrium behavior occurs: since this is a zero-probability event, Bayes' rule does not specify how the actor should update its beliefs. An arbitrary specification of beliefs in these instances can rationalize a wide variety of behavior in equilibrium, and some of that behavior can be quite odd: for instance, a resolved actor may be deterred from taking an action that would signal its strength because its opponent threatens to interpret any such behavior as evidence of weakness. The general approach to this issue has been to define increasingly stringent notions of what actors are permitted to believe when a zero-probability event occurs (Cho and Kreps 1987; Banks and Sobel 1987), but the deeper problem is that it remains unclear why an actor who observes evidence that flatly contradicts its conjecture about the world would continue to believe the model (Blume et al. 1991).

The theory also provides little guidance as to what to do when the model admits multiple solutions for the same set of parameters. The problem of multiple equilibria is unresolved, and may remain so (Cooper 1999). While the indeterminacy may be a serious hindrance for empirical testing of theoretic implications, the quest for a unique solution may be contrary to what the rich empirical reality may offer (Harsanyi and Selten 1988). Evolutionary models hold out the promise to provide a selection mechanism that eliminates some equilibria in favor of others.

Finally, game-theoretic models may quickly become intractable when multiple actors are introduced, especially when diverse players are modeled in incomplete information settings. For instance, it is not clear whether different agents should interpret observable behavior in the same way. This problem is particularly glaring for applications to international relations where most non-formal thought concerns behavior in multiplayer environments (Wagner 1986). Despite all these problems, the history of game theory suggests that it would be premature to abandon the tool, especially in the absence of a viable alternative. If anything, the development of game theory has been driven precisely by the realization of its limitations and attempts to overcome them.

References

Aumann, R.J. (1987) Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica* 55, 1–18.

Aumann, R.J., and Brandenburger, A. (1995) Epistemic Conditions for Nash Equilibria. *Econometrica* 63, 1161–80.

Axelrod, R. (1981) The Emergence of Cooperation Among Egoists. *American Political Science Review* 75, 306–18.

Axelrod, R. (1984) *The Evolution of Cooperation*. New York: Basic Books.

Baldwin, D. (ed.) (1993) *Neorealism and Neoliberalism*. New York: Columbia University Press.

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

Banks, J., and Sobel, J. (1987) Equilibrium Selection in Signaling Games. *Econometrica* 55, 647–661.

Bernheim, D. (1984) Rationalizable Strategic Behavior. *Econometrica* 52, 1007–28.

Bichieri, C. (1988) Strategic Behavior and Counterfactuals. *Synthese* 76, 135–69.

Binmore, K. (1987) Modeling Rational Players, I. *Economics and Philosophy* 3, 9–55.

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

Blume, L., Brandenburger, A., and Dekel, E. (1991) Lexicographic Probabilities and Equilibrium Refinements. *Econometrica* 59, 81–98.
Brams, S.J. (1975) *Game Theory and Politics*. New York: Free Press.
Brams, S.J. (1994) *Theory of Moves*. Cambridge: Cambridge University Press.
Brams, S.J., and Wittman, D. (1981) Nonmyopic Equilibria in 2 × 2 Games. *Conflict Management and Peace Science* 6, 39–62.
Brito, D.L. (1972) A Dynamic Model of an Armaments Race. *International Economic Review* 13, 259–75.
Bueno de Mesquita, B. (1981) *The War Trap*. New Haven: Yale University Press.
Bueno de Mesquita, B., and Lalman, D. (1986) Reason and War. *American Political Science Review* 80, 1113–29.
Bueno de Mesquita, B., and Lalman, D. (1992) *War and Reason*. New Haven: Yale University Press.
Bueno de Mesquita, B., and Morrow, J. (1999) Sorting Through the Wealth of Notions. *International Security* 24, 56–73.
Bueno de Mesquita, B., Morrow, J., Siverson, R., and Smith, A. (2003) *The Logic of Political Survival*. Cambridge: MIT Press.
Cederman, L. (1994) Emergent Polarity: Analyzing State Formation and Power Politics. *International Studies Quarterly* 38, 501–33.
Cho, I., and Kreps, D. (1987) Signaling Games and Stable Equilibria. *Quarterly Journal of Economics* 102, 179–221.
Cooper, R. (1999) *Coordination Games*. Cambridge: Cambridge University Press.
Danilovic, V. (2001) Conceptual and Selection Bias Issues in Deterrence. *Journal of Conflict Resolution* 45, 97–125.
Danilovic, V. (2002) *When the Stakes Are High: Deterrence and Conflict among Major Powers*. Ann Arbor: University of Michigan Press.
Doran, C., and Parsons, W. (1980) War and the Cycle of Relative Power. *American Political Science Review* 74, 947–65.
Dougherty, J., and Pfaltzgraff, R. (1971) *Contending Theories of International Relations*. Philadelphia: J.P. Lippincott.
Downs, G., and Rocke, D. (1994) Conflict, Agency, and Gambling for Resurrection: The Principal–Agent Problem Goes to War. *American Journal of Political Science* 38, 362–80.
Ellsberg, D. (1959) The Theory and Practice of Blackmail. Lecture at the Lowell Institute, Boston, MA, 10 Mar. Reprinted in Young (1975).
Fearon, J. (1994a) Domestic Political Audiences and the Escalation of International Disputes. *American Political Science Review* 88, 577–92.
Fearon, J. (1994b) Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model. *Journal of Conflict Resolution* 38, 236–69.
Fearon, J. (1995) Rationalist Explanations for War. *International Organization* 49, 379–414.
Fearon, J. (1997) Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs. *Journal of Conflict Resolution* 41, 68–90.
Filson, D., and Werner, S. (2002) A Bargaining Model of War and Peace: Anticipating the Onset, Duration, and the Outcome of War. *American Journal of Political Science* 46, 819–38.
Fraser, N., and Hipel, K. (1979) Solving Complex Conflicts. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-9, 805–16.
Freedman, L. (1989) *The Evolution of Nuclear Strategy*, 2nd edn. New York: St. Martin's.
Fudenberg, D., and Levine, D. (1998) *The Theory of Learning in Games*. Cambridge: MIT Press.
Grieco, J. (1988) Anarchy and the Limits of Cooperation. *International Organization* 42, 485–507.
Harsanyi, J.C. (1965) Game Theory and the Analysis of International Conflict. *Australian Journal of Politics and History* 11, 292–304.
Harsanyi, J.C. (1967) Games with Incomplete Information Played by "Bayesian" Players. Part I: The Basic Model. *Management Science* 14, 159–82.
Harsanyi, J.C. (1968a) Part II: Bayesian Equilibrium Points. *Management Science* 14, 320–4.
Harsanyi, J.C. (1968b) Part III: The Basic Probability Distribution of the Game. *Management Science* 14, 486–502.
Harsanyi, J., and Selten, R. (1988) *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press.

Haywood, O.J., Jr. (1954) Military Decision and Game Theory. *Journal of the Operations Research Society of America* 2, 65–85.
Hollist, W.L. (1977) An Analysis of Arms Processes in the United States and the Soviet Union. *International Studies Quarterly* 21, 503–28.
Howard, N. (1971) *Paradoxes of Rationality: Theory of Metagames and Political Behavior.* Cambridge: MIT Press.
Huth, P. (1999) Deterrence and International Conflict: Empirical Findings and Theoretical Debates. *Annual Review of Political Science* 2, 61–84.
Iida, K. (1993) When and How Do Domestic Constraints Matter? Two-Level Games with Uncertainty. *Journal of Conflict Resolution* 37, 403–26.
Jervis, R. (1970) *The Logic of Images in International Relations.* New York: Columbia University Press.
Jervis, R. (1972) Bargaining and Bargaining Tactics. In J.R. Pennock and J.W. Chapman, (eds.), *Coercion.* Chicago: Aldine, pp. 272–88.
Jervis, R. (1979) Deterrence Theory Revisited. *World Politics* 31, 289–324.
Jervis, R. (1988) Realism, Game Theory, and Cooperation. *World Politics* 40, 317–49.
Kadera, K. (2001) *The Power–Conflict Story: A Dynamic Model of Interstate Rivalry.* Ann Arbor: University of Michigan Press.
Kahn, H. (1960) *On Thermonuclear War.* Princeton: Princeton University Press.
Kahn, H. (1962) *Thinking About the Unthinkable.* New York: Horizon Press.
Kahn, H. (1965) *On Escalation.* New York: Praeger.
Kaplan, M.A. (1957). *System and Process in International Politics.* New York: Wiley.
Kecskemeti, P. (1958) *Strategic Surrender.* Stanford: Stanford University Press.
Keohane, R. (1984) *After Hegemony.* Princeton: Princeton University Press.
Kuhn, H. (ed.) (1997) *Classics in Game Theory.* Princeton: Princeton University Press.
Kurizaki, S. (2007) Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy. *American Political Science Review* 101, 543–58.
Leventoglu, B., and Slantchev, B.L. (2007) The Armed Peace: A Punctuated-Equilibrium Theory of War. *American Journal of Political Science* 51, 755–71.
Leventoglu, B., and Tarar, A. (2008) War and Incomplete Information. *International Studies Quarterly* 52, 533–53.
Luce, R.D., and Raiffa, H. (1957) *Games and Decisions.* New York: Wiley.
McDonald, J. (1950) *Strategy in Poker, Business and War.* New York: Norton.
McDonald, J., and Tukey, J. (1949) Colonel Blotto: A Problem of Military Strategy. *Fortune*, June.
McGinnis, M.D. (1991) Richardson, Rationality, and Restrictive Models of Arms Races. *Journal of Conflict Resolution* 35, 443–73.
Majeski, S.J., and Jones, D.L. (1981) Arms Race Modeling: Causality Analysis and Model Specification. *Journal of Conflict Resolution* 25, 259–88.
Martin, L. (1999) The Contributions of Rational Choice: A Defense of Pluralism. *International Security* 24, 74–83.
Mearsheimer, J. (1990) Back to the Future: Instability in Europe after the Cold War. *International Security* 15, 5–56.
Milner, H.V. (1997) *Interests, Institutions, and Information.* Princeton: Princeton University Press.
Mo, J. (1995) Domestic Institutions and International Bargaining: The Role of Agent Veto in Two-Level Games. *American Political Science Review* 89, 914–24.
Morgenstern, O. (1959) *The Question of National Defense.* New York: Random House.
Morgenstern, O. (1961a) The Cold War Is Cold Poker. *New York Times Magazine*, Feb. 5.
Morgenstern, O. (1961b) Review of *The Strategy of Conflict. Southern Economic Journal* 28, 105.
Morgenthau, H. (1948) *Politics Among Nations: The Struggle for Power and Peace.* New York: Knopf.
Morrow, J.D. (1989) Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining. *American Journal of Political Science* 33, 941–72.
Morrow, J.D. (1991) Electoral and Congressional Incentives and Arms Control. *Journal of Conflict Resolution* 35, 245–65.
Nalebuff, B. (1991) Rational Deterrence in an Imperfect World. *World Politics* 43, 313–35.
Nash, J. (1951) Non-cooperative Games. *Annals of Mathematics* 54, 286–95.
Niou, E., and Ordeshook, P. (1999) The Return of the Luddites. *International Security* 24, 84–96.

O'Neill, B. (1994a) Sources in Game Theory for International Relations Specialists. In M. Intriligator and U. Luterbacher (eds.) *Cooperative Models in International Relations Research*. Boston: Kluwer, pp. 9–30.
O'Neill, B. (1994b) Game Theory Models of Peace and War. In R.J. Aumann and S. Hart (eds.) *Handbook of Game Theory*, vol. 2. Amsterdam: Elsevier, pp. 995–1053.
O'Neill, B. (1999) *Honor, Symbols, and War*. Ann Arbor: University of Michigan Press.
Organski, A.F.K. (1968) *World Politics*, 2nd edn. New York: Knopf.
Organski, A.F.K., and Kugler, J. (1980) *The War Ledger*. Chicago: University of Chicago Press.
Ostrom, C.W., and Marra, R.F. (1986) US Defense Spending and the Soviet Estimate. *American Political Science Review* 80, 819–42.
Oye, K. (ed.) (1986) *Cooperation under Anarchy*. Princeton: Princeton University Press.
Pearce, D. (1984) Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica* 52, 1029–50.
Powell, R. (1987) Crisis Bargaining, Escalation, and MAD. *American Political Science Review* 81, 717–35.
Powell, R. (1988) Nuclear Brinkmanship with Two-Sided Incomplete Information. *American Political Science Review* 82, 155–78.
Powell, R. (1990) *Nuclear Deterrence Theory: The Search for Credibility*. New York: Cambridge University Press.
Powell, R. (1991) Absolute and Relative Gains in International Relations Theory. *American Political Science Review* 85, 1303–20.
Powell, R. (1996) Stability and the Distribution of Power. *World Politics* 48, 239–67.
Powell, R. (1999) The Modeling Enterprise and Security Studies. *International Security* 24, 97–106.
Powell, R. (2002) Bargaining Theory and International Conflict. *Annual Review of Political Science* 5, 1–30.
Powell, R. (2003) Nuclear Deterrence Theory, Nuclear Proliferation, and National Missile Defense. *International Security* 27, 86–118.
Powell, R. (2004) Bargaining and Learning while Fighting. *American Journal of Political Science* 48, 344–61.
Powell, R. (2007a) Defending against Terrorist Attacks with Limited Resources. *American Political Science Review* 101, 527–41.
Powell, R. (2007b) Allocating Defensive Resources with Private Information about Vulnerability. *American Political Science Review* 101, 799–809.
Putnam, R.D. (1988) Diplomacy and Domestic Politics: The Logic of Two-Level Games. *International Organization* 42, 427–60.
Quackenbush, S. (2006) National Missile Defense and Deterrence. *Political Research Quarterly* 59, 533–41.
Rapoport, A. (1964) *Strategy and Conscience*. New York: Harper and Row.
Rapoport, A. (1967) Escape from Paradox. *Scientific American* 217, 50–6.
Rapoport, A. (1968) Chicken á la Kahn. *Virginia Quarterly Review* 41, 370–89.
Richardson, L.F. (1960) *Arms and Insecurity*. Chicago: Homewood.
Riker, W.H. (1962) *The Theory of Political Coalitions*. New Haven: Yale University Press.
Riker, W.H., and Ordeshook, P. (1973) *An Introduction to Positive Political Theory*. Englewood Cliffs: Prentice Hall.
Sartori, A. (2005) *Deterrence by Diplomacy*. Princeton: Princeton University Press.
Schelling, T. (1960) *The Strategy of Conflict*. Cambridge: Harvard University Press.
Schelling, T. (1966) *Arms and Influence*. New Haven: Yale University Press.
Schultz, K. (1998) Domestic Opposition and Signaling in International Crises. *American Political Science Review* 92, 829–44.
Schultz, K. (2001) Looking for Audience Costs. *Journal of Conflict Resolution* 45, 32–60.
Selten, R. (1975) A Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games. *International Journal of Game Theory* 4, 25–55.
Senese, P., and Quackenbush, S. (2003) Sowing the Seeds of Conflict: The Effect of Dispute Settlements on Durations of Peace. *Journal of Politics* 65, 696–717.
Slantchev, B.L. (2003) The Power to Hurt: Costly Conflict with Completely Informed States. *American Political Science Review* 47, 123–33.
Slantchev, B.L. (2005) Military Coercion in Interstate Crises. *American Political Science Review* 99, 533–47.

Slantchev, B.L. (2006) Politicians, the Media, and Domestic Audience Costs. *International Studies Quarterly* 50, 445–77.
Slantchev, B.L. (2007) Feigning Weakness. Manuscript, Department of Political Science, University of California – San Diego.
Smith, A. (1998) International Crises and Domestic Politics. *American Political Science Review* 92, 623–38.
Smith, A., and Stam, A. (2004) Bargaining and the Nature of War. *Journal of Conflict Resolution* 48, 783–813.
Snidal, D. (1991) Relative Gains and the Pattern of International Cooperation. *American Political Science Review* 85:701–26.
Snidal, D. (2002) Rational Choice and International Relations. In W. Carlsnaes, T. Riesse, and B. Simmons (eds.) *Handbook of International Relations*. London: Sage, pp. 73–94.
Snyder, G. (1961) *Deterrence and Defense: Toward a Theory of National Security*. Princeton: Princeton University Press.
Snyder, G. (1971) "Prisoner's Dilemma" and "Chicken" Models in International Politics, *International Studies Quarterly* 15, 66–103.
Snyder, G., and Diesing, P. (1977) *Conflict Among Nations*. Princeton: Princeton University Press.
Stein, A. (1983) Coordination and Collaboration: Regimes in an Anarchic World. In S.D. Krasner (ed.) *International Regimes*. Ithaca: Cornell University Press, pp. 115–40.
Tarar, A. (2001) International Bargaining with Two-Sided Domestic Constraints. *Journal of Conflict Resolution* 45, 320–40.
Taylor, M. (1976) *Anarchy and Cooperation*. London: Wiley.
Trachtenberg, M. (1990/1) The Meaning of Mobilization in 1914. *International Security* 15, 120–50.
Trachtenberg, M. (1991) *History and Strategy*. Princeton: Princeton University Press.
Von Neumann, J., and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
Wagner, R.H. (1983) The Theory of Games and the Problem of International Cooperation. *American Political Science Review* 77, 330–46.
Wagner, R.H. (1986) The Theory of Games and the Balance of Power. *World Politics* 38, 546–76.
Wagner, R.H. (2000) Bargaining and War. *American Journal of Political Science* 44, 469–84.
Walt, S. (1999) Rigor or Rigor Mortis? Rational Choice and Security Studies. *International Security* 23, 5–48.
Ward, M.D. (1984) Differential Paths to Parity: A Study of the Contemporary Arms Race. *American Political Science Review* 78, 297–317.
Williams, J.D. (1954) *The Compleat Strategyst*. Santa Monica: Rand.
Wittman, D. (1979) How a War Ends: A Rational Model Approach. *Journal of Conflict Resolution* 23, 743–63.
Wright, Q. (1965) *A Study of War*. Chicago: University of Chicago Press.
Young, O. (ed.) (1975) *Bargaining*. Urbana: University of Illinois Press.
Zagare, F.C. (1984) *Game Theory: Concepts and Applications*. Beverly Hills: Sage.
Zagare, F.C. (1987) *The Dynamics of Deterrence*. Chicago: University of Chicago Press.
Zagare, F.C. (1990a) Rationality and Deterrence. *World Politics* 42, 238–60.
Zagare, F.C. (1990b) Rational Choice Models and International Relations Research. *International Interactions* 15, 197–201.
Zagare, F.C. (1999) All Mortis, No Rigor. *International Security* 24, 107–14.
Zagare, F.C. (2004) Reconciling Rationality with Deterrence: A Re-examination of the Logical Foundations of Deterrence Theory. *Journal of Theoretical Politics* 16, 107–41.
Zagare, F.C., and Kilgour, D.M. (2000) *Perfect Deterrence*. Cambridge: Cambridge University Press.

Online Resources

Game Theory.Net. At www.gametheory.net, accessed Mar. 19, 2009. A comprehensive site that has material for both the beginning and the advanced student of game theory – perhaps the best such site on the Internet. The links for educators include teaching materials, lecture notes, books and articles; and for students there is a dictionary of terms, selected applications, and course reviews. There are also links to music and movies that make use of game-theoretic concepts.

Game Theory Website v2.0. At www.holycross.edu/departments/biology/kprestwi/behavior/ESS/ESS_index_frmset.html, accessed Mar. 19, 2009. This site, which is designed for undergraduates, provides an introduction to evolutionary game theory. Although primarily of interest to those with an interest in animal behavior, it provides a fascinating exposition of the Hawk versus Dove game that may be of interest to some students of interstate conflict.

Gambit. At http://gambit.sourceforge.net, accessed Mar. 19, 2009. Gambit is a software program that can be used to construct and solve finite extensive form and strategic form games. It can be downloaded from this website. Tutorials for using the program are also available here.

Game Theory and You. At www.columbia.edu/~de11/gamethry.html, accessed Mar. 19, 2009. Syllabi from political science and economics courses that relate to game theory and politics.

David Levine's Economic and Game Theory. At http://levine.sscnet.ucla.edu/, accessed Mar. 19, 2009. Contains a general introduction to the modern theory of games with interesting examples, suggestions for further reading, and reviews of a number of game theory texts.

Al Roth's Game Theory, Experimental Economics, and Market Design Page. At http://kuznets.fas.harvard.edu/~aroth/alroth.html, accessed Mar. 19, 2009. Large archive with notes, papers, and links to other sites using game theory and experimental economics. Maintained by Al Roth.

Game Theory and Information. At http://econpapers.repec.org/paper/wpawuwpga/, accessed Mar. 19, 2009. Archive with working papers on current game-theoretic and economics research.

Ariel Rubinstein's website. At http://arielrubinstein.tau.ac.il/, accessed Mar. 19, 2009. Free downloads of Rubinstein's books, including *Bargaining and Markets* and *Modeling Bounded Rationality*.

About the Authors

Frank C. Zagare is UB Distinguished Professor of Political Science at the University at Buffalo. A former Vice President of the International Studies Association, he is the author, co-author, or editor of five books including *Perfect Deterrence* (with D. Marc Kilgour), *The Dynamics of Deterrence*, and *Game Theory: Concepts and Applications*. In 2005 he received the Susan Strange Award from the International Studies Association.

Branislav L. Slantchev is Associate Professor of Political Science at the University of California – San Diego. He studies crisis bargaining, the conduct and termination of wars, and militarized coercion. His work has appeared in the *American Political Science Review*, *American Journal of Political Science*, *International Studies Quarterly*, *Security Studies*, and *Political Research Quarterly*.