CHAPTER 4

CONTENTS

Game Theory

Frank C. Zagare

Abstract

This chapter describes the basic assumptions, and illustrates the major concepts, of game theory using examples drawn from the security studies literature. For instance, an arms race game is used to illustrate the strategic form of a game, the meaning of an equilibrium outcome, and the definition of a dominant strategy. Backwards induction and the definition of subgame perfection are explained in the context of an explication of an extensive-form game that features threats. A short review of the many applications of game theory in international politics is provided. Finally, the chapter concludes with a discussion of the usefulness of game theory in generating insights about deterrence.

Introduction

Game theory is the science of interactive decision-making. It was created in one fell swoop with the publication of John Von Neumann and Oskar Morgenstern's *Theory of Games and Economic Behavior* (1944) by Princeton University Press. Widely hailed when it was published, the book became an instant classic. Its impact was enormous. Almost immediately, game theory began to penetrate economics – as one might well expect. But soon afterwards,

applications, extensions and modifications of the framework presented by Von Neumann and Morgenstern began to appear in other fields, including sociology, psychology, anthropology and, through political science, International Relations and security studies.

In retrospect, the ready home that game theory found in the field of security studies is not very surprising. Much of the gestalt of game theory may easily be discerned in the corpus of diplomatic history and in the work of the most prominent theorists of international politics.[1] And its key concepts have obvious real-world analogues in the international arena.

Primitive concepts

The basic concept is that of a game itself. A *game* may be thought of as any situation in which an outcome depends on the choices of two or more decision-makers. The term is somewhat unfortunate. Games are sometimes thought of as lighthearted diversions. But in game theory the term is not so restricted. For instance, most if not all interstate conflicts qualify as very serious games.

In game theory, decision-makers are called *players*. Players may be individuals or groups of individuals who in some sense operate as a coherent unit. Presidents, prime ministers, kings and queens, dictators, foreign secretaries and so on can therefore sometimes be considered as players in a game. But so can the states in whose name they make foreign policy decisions. It is even possible to consider a coalition of two or more states as a player. For example, in their analysis of the July crisis of 1914, Snyder and Diesing (1977) use elementary game theory to examine the interaction between 'Russia–France' and 'Austria–Germany'.

The decisions that players make eventually lead to an *outcome*. In game theory, an outcome can be just about anything. Thus, the empirical content associated with an outcome will vary with the game being analysed. Sometimes, generic terms such as 'compromise' or 'conflict' are used to portray outcomes. At other times, the descriptors are much more specific. Snyder and Diesing use the label 'Control of Serbia' by Austria–Germany to partially describe one potential outcome of the July crisis.

Reflecting perhaps the intensity of the Cold War period in the USA in the early 1950s, almost all of the early applications of game theory in the field of security studies analysed interstate conflicts as *zero-sum games*. A zero-sum game is any game in which the interests of the players are diametrically opposed. Examples of this genre include an analysis of two World War II battles by A.G. Haywood (1954) and a study of military strategy by McDonald and Tukey (1949).

By contrast, a non-zero-sum game is an interactive situation in which the players have mixed motives; that is, in addition to conflicting interests, they may also have some interests in common. Two states locked in an economic conflict, for instance, obviously have an interest in securing the best possible terms of trade. At the same time, they both may also want to avoid the costs

associated with a trade war. It is clear that in such instances, the interests of the two states are not diametrically opposed.

The use of non-zero-sum games became the standard form of analysis in international politics towards the end of the 1950s, due in no small part to the scholarship of Thomas Schelling (1960, 1966) whose works are seminal. When Schelling's book *The Strategy of Conflict* was republished in 1980 by Harvard University Press he remarked in a new Preface that the idea that conflict and common interest were not mutually exclusive, so obvious to him, was among the book's most important contributions. In 2005, Schelling was awarded the Nobel Prize in economics for his work on game theory and interstate conflict. The award was well deserved.

Most studies also make use of the tools and concepts of non-cooperative game theory. A *non-cooperative game* is any game in which the players are unable to irrevocably commit themselves to a particular course of action. By contrast, binding agreements are possible in a *cooperative game*. Since it is commonly understood that the international system lacks an overarching authority that can enforce commitments or agreements, it should come as no surprise that non-cooperative game theory holds a particular attraction for theorists of interstate conflict.

Strategic-form games and Nash equilibria

Game theorists have developed a number of distinct ways to represent a game's structure. Initially, the *strategic-form* (sometimes called the *normal-* or the *matrix-*form) was the device of choice. In the strategic-form, players select *strategies* simultaneously, before the actual play of the game. A strategy is defined as a complete contingency plan that specifies a player's choice at every situation that might arise in a game. Figure 4.1 depicts a typical arms race game between two states, State A and State B, in strategic-form.[2] Although the generic name for this game is Prisoners' Dilemma, it is referred to here as the Arms Race game.[3]

In this representation, each state has two strategies: to *cooperate* (C) by not arming, and to *defect* from cooperation (D) by arming. If neither arm, the outcome is a compromise: a military balance is maintained, but at little cost. If both arm, both lose, as an arms race takes place, the balance is maintained, but this time at considerable cost. Finally, if one state arms and the other does not, the state that arms gains a strategic advantage, and the state that chooses not to arm is put at a military disadvantage.

Each cell of the matrix contains an ordered pair of numbers below the names of the outcomes. The numbers represent the payoff which the row (State A) and the column player (State B) receives, respectively, when that outcome obtains in a game. Payoffs are measured by a *utility* scale. Sometimes, as in this chapter, only *ordinal utilities* are, or need be, assumed. Ordinal utilities convey information about a player's relative ranking of the outcomes. In many studies of interstate conflict, however, *cardinal utilities* are assumed. A cardinal scale indicates both rank and intensity of preference.

		State B: Not arm (C)	State B: Arm (D)
State A	Not arm (C)	*Tacit arms control* (3,3)	*B gains advantage* (1,4)
	Arm (D)	*A gains advantage* (4,1)	*Arms race* (2,2)*

Key: (x,y) = payoff to State A, payoff to State B
* = Nash equilibrium

Figure 4.1 Arms Race game (Prisoners' Dilemma)

In this example, the outcomes are ranked from best (i.e. '4') to worst (i.e. '1'). Thus, the ordered pair (4,1) beneath the outcome *A gains advantage* signifies that this outcome is best for State A and worst for State B. Similarly, the outcome *Tacit arms control* is next best for both players.

In game theory the players are assumed to be instrumentally *rational*. Rational players are those who maximize their utility. Utility, though, is a subjective concept. It indicates the worth of an outcome *to a particular player*. Since different players may evaluate the same outcome differently, the rationality assumption is simply another way of saying that the players are purposeful, that they are pursuing goals (or interests) that they themselves define.

Rationality, however, does not require that the players are necessarily intelligent in setting their goals. It may sometimes be the case that the players are woefully misinformed about the world and, as a consequence, have totally unreasonable objectives. Still, as long as they are purposeful and act to bring about their goals, they may be said to be instrumentally rational.[4]

Rationality also does not imply that the players will do well and obtain their stated objective, as is easily demonstrated by identifying the *solution* to the Arms Race game. A solution to any strategic-form game consists of the identification of (1) the best, or optimal, strategy for each player, and (2) the likely outcome of the game. The Arms Race game has a straightforward solution.

Notice first that each player (State) in the Arms Race game has a *strictly dominant strategy*; that is, a strategy that is always best regardless of the strategy selected by the other player. For instance, if State B chooses not to arm, State A will bring about its next-best outcome (3) if it also chooses not to arm, but

will receive its best outcome (4) if it chooses to arm. Thus, when State B chooses (C), State A does better by choosing (D). Similarly, if State B chooses to arm, State A will bring about its worst outcome (1) if it chooses not to arm, but will receive its next-worst outcome (2) if it chooses to arm. Again, when State B chooses (D), State A does better by choosing (D). Regardless of what strategy State B selects, therefore, State A should choose (D) and arm. By symmetry, State B should also choose to defect by arming. And, when both players choose their unconditionally best strategy, the outcome is an arms race – which is next worst for both players.

The strategy pair (D,D) associated with the outcome labelled *Arms Race* has a very important property that qualifies it to be part of the solution to the game of Figure 4.1. It is called a *Nash equilibrium* – named after John Nash, the subject of the film *A Beautiful Mind* and a co-recipient of the Nobel Prize in economics in 1994 which, not coincidentally, was the fiftieth anniversary of the publication of Von Neumann and Morgenstern's monumental opus. If a strategy pair is a Nash equilibrium, neither player has an incentive to switch to another strategy, provided that the other player does not also switch to another strategy.

To illustrate, observe that if both States A and B choose to arm (D), State A's payoff will be its second best (2). But if it then decides to not arm (C), its payoff is its worst (1). In consequence, State A has no incentive to switch strategies if both states choose to arm. The same is true of State B. The strategy pair (D,D), therefore, is said to be stable or in equilibrium.

There is no other strategy pair with this property in the Arms Race game, as is easily demonstrated. For instance, consider the strategy pair (C,C) associated with the outcome *Tacit arms control*. This outcome is second-best for both players. Nonetheless, both players have an incentive to switch, unilaterally, to another strategy in order to bring about a better outcome. State B, for instance, can bring about its best outcome (4) by simply switching to its (D) strategy. Thus, the payoff pair (C,C) is not a Nash equilibrium. The same is true for the remaining two strategy pairs in this game, (C,D) and (D,C).

For reasons that will be more fully explained below, strategy pairs that form a Nash equilibrium provide a *minimum* definition of rational choice in a game. By contrast, strategy pairs that are not in equilibrium are simply inconsistent with rational choice and purposeful action. This is why only Nash equilibria can be part of a game's solution.

But notice that *both* players do worse when they are rational and select (D) than when *both* make an irrational choice and select (C). In other words, two rational players do worse in this game than two irrational players! Paradoxically, however, it is also true that *each* player always does best by choosing (D), all of which raises a very important question for the two states in our game. Can they, if they are rational, avoid an arms race and, if so, under what conditions? More generally, can two or more states ruthlessly pursuing their own interests find a way to cooperate in an anarchic international system?

Space considerations preclude an answer, game-theoretic or otherwise, to this question here. Suffice it to say that it is an issue that lies at the heart of the

ongoing debate between realists and liberals about the very nature of international politics. That the (Prisoners' Dilemma) game in Figure 4.1 both highlights and neatly encapsulates such a core problem must be counted among game theory's many contributions to the field of security studies.[5]

Even though rational players do not fare well in this game, the game itself has a well-defined solution that helps to explain, *inter alia*, why great states sometimes engage in senseless and costly arms competitions that leave them no more secure than they would have been if they had chosen not to arm. The solution is well defined because there is only one outcome in the game that is consistent with rational contingent decision-making by all of the players, the unique Nash equilibrium (D,D).

Not all games, however, have a solution that is so clear-cut. Consider, for example, the two-person game in Figure 4.2 that was originally analysed by John Harsanyi (1977), another 1994 Nobel Prize laureate in economics. As before, the two players, States A and B, have two strategies: either to cooperate (C) or to defect (D) from cooperation. State A's strategies are listed as the rows of the matrix, while B's strategies are given by the columns. Since each player has two strategies, there are $2 \times 2 = 4$ possible strategy combinations and four possible outcomes. The payoffs to State A and State B, respectively, are again represented by an ordered pair in each cell of the matrix.

Of these four strategy combinations, two are Nash equilibria, as indicated by the asterisks (*). Strategy pair (D,D) is in equilibrium since either player would do worse by switching, unilaterally, to its other strategy. Specifically, were State A to switch from its (D) strategy to its (C) strategy, which would induce Outcome CD, State A's payoff would go from '2' – A's best – to '1' – its next

		State B	
		Cooperate (C)	Defect (D)
State A	Cooperate (C)	*Outcome CC* (1,3)*	*Outcome CD* (1,3)
	Defect (D)	*Outcome DC* (0,0)	*Outcome DD* (2,2)*

Key: (x,y) = payoff to State A, payoff to State B
* = Nash equilibrium

Figure 4.2 Strategic-form game with two Nash equilibria (Harsanyi's game)

best. And if State B were to switch to its (C) strategy, B's payoff would go from '2' – its next best – to '0' – its worst. Thus, neither player benefits by switching unilaterally to another strategy, so (D,D) is a Nash equilibrium. For similar reasons, strategy pair (C,C) is also a Nash equilibrium; neither player benefits by switching, unilaterally, to its (D) strategy. By contrast, neither of the remaining two strategy pairs is stable in the sense of Nash because at least one player would gain by changing to another strategy.

The existence of two or more Nash equilibria in a strategic-form game can confound analysis. When only one Nash equilibrium exists in a game, it is easy to specify a game's solution. But when two or more equilibria exist, it is clearly more difficult to identify the likely outcome of a game or the best strategy of the players – unless there are criteria that allow discrimination among equilibria and the elimination of some stable strategy pairs from the solution set.

Of course, the possible existence of multiple Nash equilibria in a strategic-form game would not be problematic if all equilibria were *equivalent* – that is, if all extant equilibria have exactly the same consequences for the players – and *interchangeable* – in the sense that every possible combination of equilibrium strategies are also in equilibrium.

John Nash (1951) proved long ago that when multiple equilibria exist in a zero-sum game, all equilibrium pairs are both equivalent and interchangeable. But this is clearly not the case in the non-zero-sum game in Figure 4.2. The two equilibria are not equivalent simply because the player's payoffs are different under each equilibrium. For instance, State A's best outcome is associated with the strategy pair (D,D); its next-best outcome with the strategy pair (C,C). The two equilibria are also not interchangeable. Although the strategy pairs (C,C) and (D,D) are in equilibrium, the pairs (C,D) and (D,C) are not. This means that the players cannot use the strategies associated with the two Nash equilibria interchangeably.

Although the two Nash equilibria in the game in Figure 4.2 are neither equivalent nor interchangeable, there is one way in which they can be distinguished. Notice that State B's defection (D) strategy *weakly dominates* its cooperation (C) strategy; that is, it provides State B with a payoff that is at least as good, and sometimes better, than its other strategy, no matter what strategy State A selects.[6] Thus, there is a good reason to expect that State B will choose (D).

Notice also that, if State B defects, State A does better by also defecting. Given that State B defects, State A will receive its highest payoff (2) by defecting, but only its second highest payoff (1) by cooperating. Since the strategy pair (D,D) is associated with State B's unconditionally best (or *dominant*) strategy, and State A's best response to B's unconditionally best strategy, one may very well argue that it, and not strategy pair (C,C), is the equilibrium that best qualifies as the solution to Harsanyi's game.

However, before this conclusion is accepted, there is one significant objection that must be considered: the fact that strategy pair (D,D) favours State A at the expense of State B. State B's payoff is clearly better under (C,C) than it is under (D,D), while it is the other way around for State A. Is there

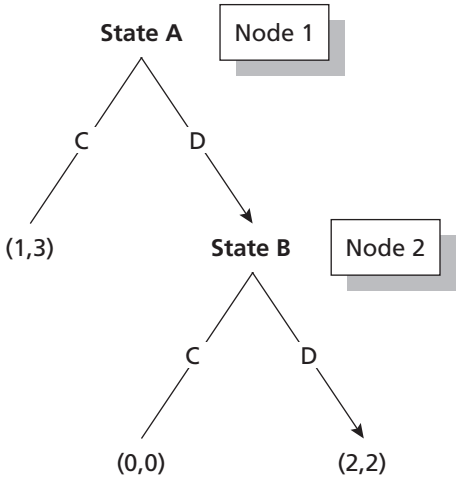
nothing that State B can do to induce the more preferred payoff associated with the equilibrium (C,C)?

One might argue that State B could do better in this game by threatening to choose (C) if State A selects (D), thereby inducing State A to choose (C) and bringing about State B's most preferred outcome. But this line of argument is deficient. To understand why, we next explore an alternative representation of Harsanyi's game, the *extensive-form*.

Extensive-form games, backwards induction and subgame perfect equilibria

Figure 4.2 represents Harsanyi's game in strategic-form; Figure 4.3 represents it in extensive-form. There are a number of important differences between the two forms of representation. In the strategic-form, players select strategies which, it will be recalled, are a complete plan of action specifying what a player will do at every decision point in a game. As well, the players are assumed to make their choice simultaneously or, in what amounts to the same thing, without information about what strategy the other player has selected.

By contrast, in the extensive-form, the players make *moves* sequentially; that is, they select from among the collection of *choices* available at any one time. In the extensive-form, moves are represented by *nodes* on a game tree. The *branches* of the tree at any one node summarize the choices available to a player at a particular point in a game. The payoffs to the players are given by an ordered pair at each terminal node. In an extensive-form game of *perfect*



Key: (x,y) = payoff to State A, payoff to State B
→ = rational choice

Figure 4.3 Extensive-form representation of Harsanyi's game

information, the players know where they are in the game tree whenever there is an opportunity to make a choice. Harsanyi's game is an example of a game of perfect information. In a game with *imperfect information*, the players may not always know what prior choices have been made.

To solve any extensive-form game, a procedure known as *backwards induction* must be used. As its name suggests, backwards induction involves working backwards up the game tree to determine, first, what a rational player would do at the last node of the tree, what the player with the previous move would do given that the player with the last move is rational, and so on until the first node of the tree is reached. We will now use this procedure to analyse the extensive-form representation of Harsanyi's game. More specifically, we now seek to establish why State B cannot rationally threaten to select (C) at node 2 in order to induce State A's cooperation at node 1, thereby bringing about State B's highest ranked outcome (1,3).

To this end, we begin by considering the calculus of State A at the first node of the tree. At node 1 State A can either select (C) and induce its second-best outcome, or select (D), which might result either in State A's best or its worst outcome. Clearly, State A should (rationally) choose (C) if it expects State B to also select (C), since the choice of (D) would then result in State A's worst outcome. Conversely, State A should select (D) if it expects State B to select (D), since this induces State A's best outcome. The question is: What should State A expect State B to do? Before we can answer this question, we must first consider State B's choice at the last node of the tree.

If State A assumes that State B is rational, then State A should expect State B to select (D) if and when State B makes its choice at node 2. The reason is straightforward: State B's worst outcome is associated with its choice of (C), its next-best outcome with its choice of (D). To expect State B to carry out the threat to choose (C) if A chooses (D), then, is to assume that State B is irrational. It follows that for State B to expect State A to select (C) is to assume that State B harbours irrational expectations about State A. To put this in a slightly different way, State B's threat is not credible; that is, it is not rational to carry out. Since it is not credible, State A may safely ignore it.

Notice what the application of backwards induction to Harsanyi's game reveals: State B's rational choice at node 2 is (D). In consequence, State A should also choose (D) at node 1. Significantly, the strategy pair (D,D) associated with these choices is in equilibrium in the same sense that the two Nash equilibria are in the strategic-form game of Figure 4.2: neither player has an incentive to switch to another strategy provided the other player does not also switch. But, also significantly, the second Nash equilibrium (C,C) is nowhere to be found. Because it was based on an incredible threat, it was eliminated by the backwards induction procedure.

The unique equilibrium pair (D,D) that emerges from an analysis of the extensive-form game of Figure 4.3 is called a *subgame perfect equilibrium*.[7] The concept of subgame perfection was developed by Reinhard Selten (1975), the third and final recipient of the 1994 Nobel Prize in economics.[8] Selten's perfectness criterion constitutes an extremely useful and important refinement

of Nash's equilibrium concept. It is a refinement because it eliminates less than perfect Nash equilibria from the set of candidates eligible for consideration as a game's solution. As well, Selten's idea of subgame perfection helps us to understand more deeply the meaning of rational choice as it applies to individuals, to groups, or even to great states involved in a conflictual relationship.

It is important to know that all subgame perfect equilibria are also Nash equilibria, but not the other way around. As demonstrated above, those Nash equilibria, such as the strategy pair (C,C) in the game in Figure 4.2, which are based on threats that lack credibility, are simply not perfect. As Harsanyi (1977: 332) puts it, these less than perfect equilibria should be considered deficient because they involve both 'irrational behavior and irrational expectations by the players about each other's behavior'.

Applications of game theory in security studies

Speaking more pragmatically, the refinement of Nash's equilibrium concept represented by the idea of a subgame perfect equilibrium and related solution concepts – such as *Bayesian Nash equilibria* and *Perfect Bayesian equilibria* – permits analysts to develop more nuanced explanations and more potent predictions of interstate conflict behaviour when applying game theory to the field of security studies.[9] It is to a brief enumeration of some of these applications, and a specific illustration of one particular application, that we turn next.

As noted earlier, applications, extensions, modifications and illustrations of game-theoretic models began to appear in the security studies literature shortly after the publication of *Theory of Games and Economic Behavior* (1944). Since then, the literature has grown exponentially and its influence on the field of security studies has been significant.[10] As Walt has observed:

> 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.
>
> (Walt 1999: 5)

Walt (1999: 7) goes on to express the fear that game-theoretic and related rational choice models are becoming so pervasive, and that their influence has been so strong, that other approaches are on the cusp of marginalization.

Although Martin (1999: 74) unquestionably demonstrates, empirically, that Walt's fear is 'unfounded', there is little doubt that game-theoretic studies are now part and parcel of the security studies literature.

Among the subject areas of security studies that have been heavily influenced by game-theoretic reasoning are the onset (Bueno de Mesquita and Lalman 1992) and escalation (Carlson 1995) of interstate conflict and war, the consequences of alliances (Smith 1995) and alignment patterns (Zagare and Kilgour 2003), the effectiveness of missile defence systems (Powell 2003, Quackenbush 2006), the impact of domestic politics on interstate conflict (Fearon 1994), the dynamics of arms races and the functioning of arms control (Brams and Kilgour 1988), the spread of terrorism (Bueno de Mesquita 2005), the dangers of nuclear proliferation (Kraig 1999), the implications of democratization for coercive diplomacy (Shultz 2001), the characteristics of crisis bargaining (Banks 1990), and the operation of balance of power politics (Niou *et al*. 1989), to name but a few.[11] In addition, as noted above, game-theoretic models have played a central role in the debate between realists and liberals about the relative importance of absolute and relative gains and about the possibility of significant great power cooperation (see note 5).

It is clear, however, that there has been no area of security studies in which game theory has been more influential than in the study of deterrence. Accordingly, I now turn to a brief discussion of this subject and attempt to illustrate, with a simple example, how game theory can help not only to clarify core concepts, but also to shed light on the conditions that lead to successful deterrence.

Although it may be somewhat of a stretch to say that Schelling was the inventor of classical deterrence theory, as does Zakaria (2001), his work is a good place to start (for an overview see Zagare 1996). Like all classical deterrence theorists, Schelling's work is characterized by two core assumptions: (1) that states (or their decision-makers) are rational; and (2) that, especially in the nuclear age, war or conflict is the worst possible outcome of any deterrence encounter. It is not difficult to demonstrate that these two assumptions are incompatible with the conclusion of most deterrence theorists that bilateral nuclear relationships, such as that between the USA and the Soviet Union during the Cold War, are inordinately stable.

To see this, consider now the Rudimentary asymmetric deterrence game as given in Figure 4.4. In this, perhaps, the simplest deterrence game one can imagine, State A begins play at node 1 by deciding whether to *concede* (C) and accept the status quo, or to *demand* (D) its alteration. If State 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*. Notice that the endpoints of this simple deterrence game list outcomes rather than player payoffs. I list outcomes and not payoffs in this example in order to use the same game-form to analyse the strategic implications of more than one payoff configuration.

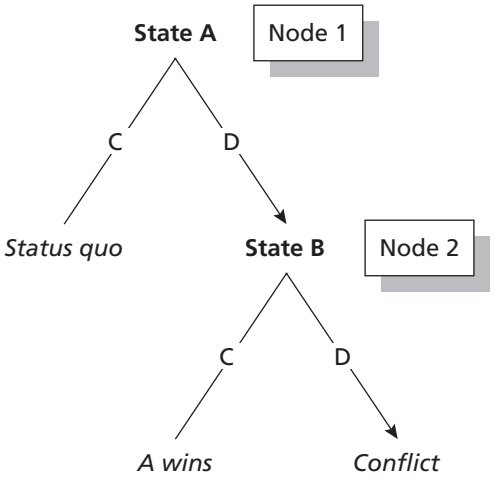


Figure 4.4 The rudimentary asymmetric deterrence game

Next we determine what rational players would do in this game – given the assumption that *Conflict* is the worst outcome for both players – by applying backwards induction to the game tree. Since the application of this procedure requires one to work backwards up the game tree, we begin by considering State B's move at decision node 2.

At node 2, State B is faced with a choice between choosing (C), which brings about outcome *A wins*, and choosing (D), which brings about *Conflict*. But if *Conflict* is assumed to be the *worst* possible outcome, State B, if it is rational, can *only* choose to concede since, by assumption, *A wins* is the more preferred outcome.

Given that State B will rationally choose to concede at node 2, what should State A do at node 1? State A can concede, in which case the outcome will be the *Status Quo*, or it can defect, in which case the outcome will be *A wins* – because a 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). Thus, given the core assumptions of classical deterrence theory, the *Status Quo* is unstable and deterrence rationally fails.

To put this in a slightly different way, one can reasonably assume that states are rational, and one can also reasonably assume that war is the worst imaginable outcome for all the players, but one cannot make both these assumptions at the same time and logically conclude, as classical deterrence theorists do, that deterrence will succeed.

Logically inconsistent theories are clearly problematic. Since *any* conclusion can be derived from them, inconsistent theories can explain *any* empirical observation. Inconsistent theories, therefore, are non-falsifiable and of little practical use. When used properly, formal structures, like game theory, can help in the identification of flawed theory.

If the core assumptions of classical deterrence theory are inconsistent with the possibility of deterrence success, what assumptions are consistent? It is easy to demonstrate that in the rudimentary asymmetric deterrence game the *Status Quo* may remain stable, and deterrence may succeed, but only if State B's threat is credible in the sense of Selten; that is, if it is rational to carry out.

To understand this, assume now that State B prefers *Conflict* to *A wins*. (Note that this assumption implies that *Conflict* is not the worst possible outcome for State B.) With this assumption, State B's rational choice at node 2 changes. Given its preference, its rational choice at node 2 is now to choose (D) and deny State A's demand for a change in the *Status Quo*.

However, State B's rational choice is not the only rational choice that changes with this new assumption. The rational choice of State A is also different. Applying backwards induction to State A's decision at node 1 now reveals a choice between *Status Quo* and *Conflict*. This means that the *Status Quo* will persist, and deterrence will succeed, *as long as State A's preference is for peace over war*. On the other hand, it will fail whenever this latter preference is reversed, even when State B's node 2 threat is credible.

At this juncture, two final observations can be made. The first is about the relationship between credible threats and deterrence success. Apparently, credibility is not, as Freedman (1989: 96) claims, the 'magic ingredient' of deterrence. As demonstrated above, a credible threat is not sufficient to ensure deterrence success. Deterrence may rationally fail even when all deterrent threats are rational to execute.

Still, in order to explain even the possibility of deterrence success in this simple example, a core assumption of classical deterrence theory had to be modified. But any analysis that proceeds from a different set of assumptions will constitute an entirely different theory. This is no small matter. As illustrated in the films *Sliding Doors* and *Run Lola Run*, and as demonstrated in Zagare (2004) and Zagare and Kilgour (2000), small differences in initial assumptions can have important theoretical consequences and significant policy differences. It is one of the strengths of game theory that its formal structure facilitates the identification of inconsistent assumptions, highlights the implications of initial assumptions, and increases the probability of logical argumentation.

Coda

This chapter provides a gentle introduction to the key concepts and assumptions of game theory as it applies to the field of security studies. The examples used to illustrate many of these terms were meant to be suggestive, and not definitive. In the space of such a short chapter, this is the best that could be done. And although an attempt has been made to point the reader to relevant applications of the theory, this effort, too, can only be thought of as being cursory. The securities studies literature that draws on, or has been influenced by, game-theoretic reasoning is vast. Nonetheless, the reader should

now possess the conceptual tools that are a prerequisite for further exploration of this increasingly important body of literature.

Notes

1 For the connections between realism and game theory, see Jervis 1988.
2 For obvious reasons, such a game is called a two-person game. Games with three or more players are referred to as *n-person games*. The latter are not discussed in this brief chapter.
3 Space considerations preclude a discussion of the story that gives this game its more common name. It is told, however, in most game theory textbooks, including Zagare 1984.
4 For an extended discussion of the rationality assumption see Zagare 1990.
5 A good place to start when exploring this and related issues is Oye 1986. Baldwin (1993) contains a useful collection of articles, many of which are seminal. Axelrod (1984), who provides one prominent game-theoretic perspective, should also be consulted. See also Chapter 10, this volume.
6 By contrast, a *strictly dominant strategy* always provides a player with a strictly higher payoff than any other strategy, no matter what strategies other players select. Both players in the Arms Race game in Figure 4.1 possess strictly dominant strategies. For a further discussion of this and related concepts, see Zagare 1984.
7 A *subgame* is that part of an extensive-form game that can be considered a game unto itself. For a more detailed definition, with pertinent examples, see Morrow 1994: ch. 2.
8 Recall that John Nash and John Harsanyi were the other two.
9 Nash and subgame perfect equilibria are the accepted measures of rational behaviour in games of *complete* information, in which each player is fully informed about the preferences of its opponent. In games of *incomplete* information in which at least one player is uncertain about the other's preferences, rational choices are associated with *Bayesian Nash equilibria* (in strategic-form games) and with *perfect Bayesian equilibria* (in extensive-form games). See Gibbons (1992) for a helpful discussion.
10 An insightful review of the accomplishments and the limitations of the approach may be found in Bueno de Mesquita (2002). See Brams (2002) for an example of the theory in action.
11 This listing is meant to be suggestive. It is by no means exhaustive. Useful reviews include O'Neill (1994a, 1994b) and Snidal (2002).

Further reading

Michael Brown, Owen R. Coté Jr., Sean M. Lynn-Jones and Steven E. Miller (eds), *Rational Choice and Security Studies* (Cambridge, MA: MIT Press,

1999). Contains a spirited debate about the contributions of game-theoretic and related approaches to the security studies literature.

Sylvia Nasar, *A Beautiful Mind* (New York: Simon & Schuster, 1998). A very readable biography of John Nash, one of the central figures of game theory.

Thomas C. Schelling, *Arms and Influence* (New Haven, CT: Yale University Press, 1966). One of the earliest and certainly one of the most influential works to draw on game theory in the analysis of interstate conflict.

Duncan Snidal, 'Rational choice and International Relations' in Walter Carlsnaes, Thomas Risse and Beth A. Simmons (eds), *Handbook of International Relations* (Thousand Oaks, CA: Sage, 2002). A fair and balanced assessment of the strengths and weaknesses of the rational choice approach.

Frank C. Zagare and D. Marc Kilgour, *Perfect Deterrence* (Cambridge: Cambridge University Press, 2000). Uses simple and advanced game-theoretic models to develop a general theory of interstate conflict initiation, limitation, escalation and resolution.