In the Shadow of Power

States and Strategies in International Politics

ROBERT POWELL

PRINCETON UNIVERSITY PRESS
PRINCETON, NEW JERSEY
threatened states tend to balance by aligning with other threatened states or bandwagon by aligning with the source of the threat? Walt (1987) and Waltz (1979) believe that states generally balance, whereas Schroeder (1994a, 1994b) concludes from his study of European politics that states bandwagon or wait while others fight more often than they balance.

Alignment decisions are extremely complex. Consider the simplest situation in which there are three states, S1, S2, and S3, and S1 is threatening to attack S2. At a minimum, S3 has three options. It can bandwagon by joining the threatening state S1, balance by aligning with S2 against the would-be attacker, or wait by not joining either S1 or S2.

Each of these alternatives has advantages and disadvantages. Waiting, at least initially, avoids the immediate cost of fighting, but it also entails a risk that the confrontation between S1 and S2 will result in an unfavorable shift in the distribution of power against S3. This latter factor encourages S3 to enter the fray. Once S3 has decided to enter the conflict, it can maximize its chances of being on the winning side by joining the stronger of the two potential coalition partners. This consideration makes S3 more likely to join the stronger state. But there is an opposing factor. If S3 aligns with the stronger state and that coalition prevails, then S3 will be in a weaker and more vulnerable position with respect to its coalition partner than it would have been had it aligned with the weaker state. How these competing pressures play themselves out and which, if any, generally dominates the others is unclear.

The analysis in chapter 5 suggests two broad conclusions. The first is a note of caution. The model used to study these issues is intended to be as simple as possible. Even so, the model shows that a large number of important assumptions must be made about the technology of coercion and, in particular, about the ways that alliances and war affect the distribution of power. Unfortunately, little empirical or theoretical evidence exists to guide us in making these assumptions. The results of the analysis are therefore tentative and must be interpreted carefully. Of course, this caution also applies to other investigations of bandwagoning. These assumptions must be made in any analysis whether formal or not. Formalization only makes them explicit. Second, the analysis does not support the claim that states generally balance. Bandwagoning and waiting are more common. But which of these three behaviors occurs depends in a complicated way on the distributions of power and benefits, the technology of coercion, and the states’ relative resolve.

### The Role of Formal Models

Chapters 2 through 5 use a series of game-theoretic models to examine the three ways that states can respond to threats, but these chapters do not analyze the games formally. Rather, the models help fix ideas and make the analysis of these responses more precise than would otherwise be possible. The discussion in these chapters emphasizes the insights and intuitions that emerge from the models about the strategic problems states face, the trade-offs these problems create, and the ways that states resolve these trade-offs. This discussion does not presume anything more than a passing familiarity with some basic game-theoretic concepts like a game tree, a strategy, and an equilibrium, which are reviewed in appendix 1.12 A technical analysis of the games is provided in appendixes at the end of the book, but the following chapters are self-contained and can be read without referring to those appendixes.

Although this study does not emphasize the technical details of models, it does pursue an approach based on a series of models, and this raises a general question about the role and usefulness of models in studying the complex problems that characterize international politics. Suppose we take the goal of international relations theory to be that of explaining and, to the extent that it is possible, making predictions about the empirical world. Given this goal, what is to be gained from looking at very simple models—models that are much too simple to capture the historic richness and detail of the actual decisions that leaders and others make? How do models, especially formal mathematical models, help explain? What are some of the relative advantages of a modeling-based approach over other kinds of approaches?

This section addresses these questions by first describing what the modeling enterprise is and how it works. This enterprise is not necessarily based on formal, mathematical models, and, indeed, most of the efforts to analyze the stylized model of the international system described above have not relied on formal models. Nevertheless, the models developed in this volume are mathematical, and this section’s second task is to discuss some of the general advantages and disadvantages of formalization. There are, however, different kinds of mathematical models, and each brings different strengths and weaknesses. The third task is to describe what game theory brings. Finally, this section concludes by

---

12Morrow (1994b) provides a comprehensive introduction to game theory.
examining some of the factors that need to be kept in mind when evaluating the game-theoretic models developed in the following chapters.

The Modeling Enterprise

The modeling enterprise is a process. It is a way of trying to use a series of simple models to understand complicated things. At its best, the enterprise consists of an iterative procedure in which research moves back and forth or iterates between a modeling realm and a more empirical realm in what Roger Myerson (1992) calls a “modeling dialogue.” Understanding grows with each round of iteration even if that understanding is simply a clearer sense of what we really do not understand very well. As Robert Jervis puts it, “Good theories do not spring full-blown from the minds of a few scholars. Rather, they develop as people test them and examine their internal dynamics and causal linkages” (1979, 303).

Although models may be very simple and highly stylized, they are generally inspired by an empirical problem. Models are a constrained, best effort to capture what the modeler believes to be the essence of a complex empirical phenomenon or at least an important aspect of it. A good model provides at least a partial picture of a causal mechanism. Indeed, the modeler may believe that several causes are at work but nevertheless designs a model to focus on only one facet of a problem. A good model need only explain how some things are related to some other things and why.13

Two kinds of constraints limit a model. The first is tractability. A model is a tool and a tool must be simple enough to use. The second constraint is the modeler's current understanding of the problem at hand. When constructing a model, the modeler tries to build in what she believes the essential cause or causes to be in order to see if they actually can explain at least part of the outcome.

Once a model has been specified, deductions can be derived from it. The specification and analysis of a model as well as the derivation of hypotheses is work conducted in the modeling realm. The dialogue then moves to the empirical realm where the deductions are evaluated or


scrutinized empirically. Data and history help discipline the model. In principle, the model might get everything right in the sense that all of the deductions are borne out empirically. In practice, this almost never happens, especially in the social sciences. At best, the model gets some important things right and some things wrong. At worst, the model seems to get everything wrong.14

What does one do with an analysis that gets some important things right and others wrong? One reading of Karl Popper (1959), which Imre Lakatos (1970, 103–163) calls naïve falsificationism, might suggest that the model should be abandoned since some of the deductions derived from it have been falsified. The modeling enterprise takes a different tack. The fact that the model got some important things right is taken as a sign that the model is tapping into and capturing important aspects of the underlying causes. This directly contributes to our understanding of those causes.

The fact that the model got some things wrong also contributes to our understanding, albeit indirectly. A model is a distillation of the modeler’s best current understanding of the causal factors underlying the phenomena being modeled. When that model gets important things wrong or, worse, everything wrong, it at least tells the modeler that his or her current understanding of the phenomena is inadequate. To the extent that the model also reflects the current understanding present in the existing literature, the model’s deficiencies also show more generally that the current understanding expressed in the literature is also inadequate. When models get things wrong, they contribute to our understanding by helping to outline the limits of our understanding and thereby motivating us to rethink the issues and search for new ideas.

New ideas can come from the modeling realm. Models, especially in the early stage of the modeling enterprise, often contain assumptions the modeler believes to be very restrictive but that greatly simplify the analysis. In these circumstances, the modeler, of course, hopes to learn something substantively interesting from the simple model. But she also expects to be better able to deal with less restrictive models after first learning how the simpler model works. Models are tools, and sometimes it helps to master a simple tool first even if the ultimate goal is to learn

14In a trivial sense, a model is certain to get some things wrong. Models make simplifications and thus incorrectly describe the thing that is being simplified. The more important issue is whether these simplifications significantly distort the analysis of the problem at hand.
how to work with a more complicated tool. Thus, when a simple model goes astray and gets things wrong, it is natural to ask if some of these restrictive assumptions can be relaxed in light of what has been learned from the simple model.

Indeed, a modeler often expects that the simple model will go astray and will want to try to relax some of these restrictions even before moving back to the empirical realm. Lakatos describes this process well:

Few theoretical scientists engaged in a research programme pay undue attention to "refutations." They have a long-term research policy that anticipates these refutations. This research policy, or order of research, is set out—in more or less detail—in the positive heuristic of the research programme... which lists a chain of ever more complicated models simulating reality... (1970, 135)

The Italian economist and sociologist, Vilfredo Pareto, put the same point more concretely.

It was a fortunate circumstance for the foundation of celestial mechanics that in Kepler's time observations of the planet Mars were not very exact. If they had been he would not have detected an ellipse in the curve traversed by that planet and so would not have discovered the laws of planetary movement. It was also fortunate that he elected to study the movements of Mars rather than those of the Moon, which is subject to much greater disturbances.

What at the time was the work of chance must now be done by the method of successive approximations. Every now and then scientific theories of economics and sociology are challenged as disregarding certain particulars. That, instead, is a merit. One must first obtain a general concept of the thing one is studying, disregarding details, which for the moment are taken as perturbations; and then come to particulars afterwards, beginning with the more important and proceeding successively towards the less important. (1935, 322–23)

Treating states as rational unitary actors is an example of a restrictive assumption that may be relaxed over time. Although the stylization of the international system described above and the models developed below assume states to be rational unitary actors, this assumption does not mean that domestic politics is unimportant. Rather, this assumption reflects, first, the hope that we can learn some interesting things by assuming that states are rational unitary actors and, second, that studying these kinds of models will also help us relax this assumption in interesting ways in subsequent work. For example, models that take states to be rational unitary actors suggest that asymmetric information is often crucial to explaining why interstate bargaining breaks down in war. In light of this unitary-actor finding, Schultz (1996) and others are trying to explain the "democratic peace" by breaking down the unitary-actor assumption and explicitly modeling some aspects of democratic domestic politics. The goal of this work is to understand how democratic institutions affect the likelihood of war by changing the informational asymmetries that exist between states during a crisis.

The need to start with simple, if restrictive, assumptions—to walk before we can run, as it were—explains why modelers sometimes do not find the criticism that this or that model is too simple to be very helpful. Modelers often think in terms of a series of models that evolve through a modeling dialogue. They already know that the current model is in some sense too simple and that making it more general is a good idea. The problem is to figure out how to do so in an interesting and fruitful, yet tractable, way.

New ideas can also come from the empirical as well as the modeling realm. In international relations, there are often only a few cases and a myriad of factors that could be at work. A detailed historical knowledge and deep sense of the cases coupled with an understanding of a model's successes and failures may suggest which factors actually are at work and should be examined more closely. Factors once thought to be irrelevant or unimportant may now take on greater significance.

Armed with these new ideas, the modeling enterprise moves back to the modeling realm where these ideas are incorporated in new or modified models. New deductions are derived, and the dialogue moves back to the empirical realm where these deductions are evaluated. Ideally these new models get more right than the previous ones. But these new models will almost surely continue to get some things wrong, and another round or iteration will begin with a search for new ideas.

15Chapter 6 discusses some of the formal work that is already being done on domestic politics.
16See Fearon (1995b) and chapter 3 below for a discussion of this.
17The "democratic peace" puzzle is used to explain the empirical observations that democratic states do not seem to fight each other even though they seem to be, on the whole, as likely to engage in war as non-democratic states. The volume of work on the problem is now quite large. For surveys and contributions to it, see Ray (1998), Rousseau et al. (1996), and Russett (1993). For other efforts to break the unitary-actor assumption, see Downs and Rocke (1994, 1995) and Bueno de Mesquita et al. (1997, 1998).
The iteration that defines the modeling enterprise occurs whether the models are mathematical constructs or employ ordinary language. Indeed, much of the existing work in international relations theory moves back and forth between non-mathematical but nevertheless theoretical models and arguments and the empirical realm of historical case studies and, less frequently, larger statistical studies. Christensen and Snyder's (1990) study of alliance dynamics exemplifies this iterative procedure.

In an effort to explain some variation in alliance behavior, they add the offense-defense balance to Waltz's (1979) theory. Based on this modification, they conclude that alliances will be tighter if there are large offensive advantages. Tighter alliances in turn imply that states run a higher risk of being dragged into a war by an ally because they are "chain-ganged" together. If, by contrast, the defense has the advantage, alliances will be looser and states are more inclined to "pass the buck" by letting others pay the costs of maintaining the alliance.

Moving to the empirical realm, Christensen and Snyder find that these conclusions turn out to be exactly backward in the two historical cases they consider. Alliances were tighter before the First World War and looser before the Second World War. But the defense had the advantage in the former, and the offense had the advantage in the latter. Given these findings, Christensen and Snyder move back to the theoretical realm where they modify their formulation by stipulating that it is the decision-makers' perception of the offense-defense balance that matters, and then they return to the empirical realm where they find that this modification does seem to fit the cases of the First and Second World Wars. This fit, however, is not terribly compelling because these cases also motivated the modification (Christensen and Snyder 1990, 145), and Morrow (1993) argues that this formulation does not account for the failure of Austria and France to ally against Prussia during the 1860s. Morrow then moves back to the theoretical realm where he draws on Altfield's (1984) model in an effort to explain alliance behavior.

In sum, the modeling enterprise is an iterative procedure or dialogue in which research moves back and forth between a more theoretical realm and a more empirical realm. Models do not drive this process; ideas about possible explanations of empirical phenomena do. However, one can often trace this iterative procedure in the literature in which one sometimes finds a series of related models where each model focuses on a new facet of the problem by trying to relax an assumption made in a previous model.

This description is, of course, an idealization of the modeling approach. It often falls short in practice because there is too little dialogue. Sometimes work stays in one realm or the other too long. Albert Einstein once observed about the relation between Euclidean geometry and experience, "In so far as geometry is certain, it says nothing about the actual world, and in so far as it says something about our experience, it is uncertain." 18 His observation about mathematics applies equally well to any logical deduction. Work that remains in the modeling realm too long can begin to mistake logical deductions for empirically established explanations and become substantively sterile. By contrast, work that remains too long in the empirical realm can begin to mistake descriptions of specific factors and taxonomies of different kinds of factors for causal explanations. 19

Some Advantages and Disadvantages of Formal Models

If the modeling enterprise, whether it be based on mathematical or non-mathematical models, iterates between a more theoretical realm comprised of models and abstractions and a more empirically oriented realm, what are some of the specific advantages that formal, mathematical models offer? Mathematical models give us "a clear and precise language for communicating insights and notions" (Kreps 1990, 6). They help us discipline our thinking about what we are trying to model. Formal models provide a kind of accounting mechanism that enables us to think through some issues more carefully than ordinary-language models can. Accounting schemes make a firm's financial situation more transparent to those both inside the firm and outside it. Formal models make arguments more transparent both to those making them and to those to whom the arguments are made.

This improved transparency comes from two sources. First, models must be fully specified or closed before they can be analyzed. Closing a

---

18 Quoted in Frank (1947, 177).
19 Of course, individuals may specialize in working primarily in one realm or the other. Mastering the skills and knowledge needed to work in either realm may require a substantial investment, and this tends to encourage specialization. If this division of labor does occur, then those specializing in more modeling-oriented work and those specializing in more empirically oriented work need to be able to communicate with each other in order to engage in a dialogue. Research as a whole needs to move back and forth, whether it be through an internal dialogue within an individual or an external dialogue among different individuals.
model often reveals that important but previously unappreciated assumptions have to be made in order to support an argument. Models help make critical assumptions more explicit. Second, the links from assumptions to conclusions are clearer in formal models. Indeed, the derivation of conclusions frequently takes the form of mathematical proofs or demonstrations. These clearer linkages in turn make it easier to trace the effects of changing one or more of the assumptions.

The benefits to more transparent arguments are at least threefold. First, greater clarity provides a check on the internal logic of existing arguments. These checks are obviously important when they reveal that an argument is wrong or, as more commonly happens in international relations theory, incomplete. And, much recent formal work in international relations theory—as well as the following chapters—shows that many widely accepted “conclusions” do not follow from the assumptions said to imply them. These arguments are generally not wrong. Indeed, they reflect a deep sense of international politics, and the assumed causes are almost certainly linked to the conclusions said to follow them. But formalization shows that other important and previously unappreciated assumptions often must be added to the arguments in order to complete them.

These additional assumptions in effect narrow the range of conditions in which the original arguments should be expected to hold. In the language of statistics, these assumptions specify what to “control” for when testing the arguments. Having an idea of what to control for may in turn make it easier to find empirical regularities in case studies as well as large data sets. Indeed, it is remarkable how few robust, well-established empirical regularities we have in international relations—the absence of war among democratic states is one of the very few—and part of the reason for this lack may be that the hypotheses we have been testing have been too broad because we have not known what to control for.

Checking the internal logic of an argument is also important even if in the end the check simply confirms the argument. If this happens, especially if simple models are used to do it, the formal analysis may seem trivial or superfluous. It may appear that the model exists for the sake of having a model—reflecting a kind of “modeling mania”—and just tells us “something we already knew.” The model is simply putting old wine in new bottles.

This charge misses an important point. We do not know that an existing argument satisfies the accounting standards formalization imposes until it has been modeled, and, as just noted, recent formal work has shown that many seemingly simple and straightforward arguments do not hold up very well when we try to model them. Some old wine turns out to be vinegar when we try to put it in new bottles.

The second benefit of formalization is that the more explicit statement of assumptions and the tighter deductive links between assumptions and conclusions facilitate the derivation of new and directly testable hypotheses. An excellent example of this is Fearon’s work (1994). He uses an asymmetric-information formulation to derive new hypotheses about the factors that make deterrence more likely to work. He then shows that these hypotheses fit the existing data better than other prevalent arguments do. More specifically, Fearon’s asymmetric-information model of a crisis leads to the hypotheses that measures of a defender’s interest in protecting a protégé which are known before any direct threat against the protégé is made—for example, the level of trade between the defender and protégé—should make threats to this protégé less likely. But these ex ante indicators should be negatively related to the likelihood that a challenger will back down after it has already made a threat. These are certainly new, non-obvious hypotheses, and they fit Huth and Russett’s (1988) data better than the existing alternatives do.

A third, broader potential benefit of formalization is that analyzing a model and then asking what accounts for the outcome sometimes leads to new insights and new ways of thinking about a large set of issues. For example, the models in the following chapters do make some specific predictions. But the models are very spare and the mechanisms they highlight are too general to explain particular outcomes in any degree of detail. Too much has been left out. But these template-models, along with the broader notions of commitment problems, asymmetric information, and the technology of coercion, provide insights and a framework for approaching a wide range of issues. These insights and the framework supporting them are most powerful and most useful when they go beyond the formal models that inspired them and help us think through complicated real-world problems which may at present be too hard to model very well.

In sum, mathematical modeling provides a language that makes it possible to define our terms more precisely and less ambiguously and to show that certain precise assumptions lead to other precise conclusions. It also allows us to stretch our analyses and to unify them; once we have worked our way through the logic that assumptions A imply conclusions X, we may see how assumptions A’ lead to conclusions X’ by the “same basic argument.” It allows us to appreciate how
critical are certain (often implicit) assumptions: If A leads to X, but a slight change in A to A' leads to not X, then we can appreciate that X or not X depends on the seemingly slight differences between A and A'; hence X is not a very robust conclusion. Taking logical deductions back to the real world, where the satisfaction of assumptions A or A' is a matter of some controversy, our developed intuition concerning what assumptions lead to which conclusions, together with a sense of how closely the real world conforms to A or A', gives us the courage to assert that X will or will not pertain with very high probability. (Kreps 1997, 63–4)

This form of reasoning is used throughout the following chapters to examine several important and very influential arguments in international relations theory.

The advantages of formalization come at a price. Requiring arguments to satisfy a different set of accounting standards may make it impossible, at least in the short run, to study some important ideas and insights because no one can figure out how to investigate them with arguments that meet the new standards. Indeed, because what we "know" is partly a function of the standards by which we evaluate arguments, imposing a different set of standards may mean that at least in the first stage, we "know" less that we thought we did. "Model-building, especially in its early stages, involves the evolution of ignorance as well as knowledge; and someone with powerful intuition, with a deep sense of the complexities of reality, may well feel that from his point of view more is lost than is gained." (Krugman 1995, 79). Whether more is lost than gained, whether truly important insights can be distinguished from those that only seem impressive, and whether the former can ultimately be incorporated in the modeling enterprise is a judgment that can be made only over the long run after much work has been done.

What is clear, however, is that many widely accepted arguments in international relations theory appear to be incomplete and in need of qualification when they are subjected to the accounting standards and greater transparency of a model. Many "conclusions" do not follow from the stated assumptions, and completing the argument sometimes reverses the conclusions or at least qualifies them by narrowing the range of circumstances in which they hold. Chapter 2, for example, shows that an anarchic environment does not imply a concern for relative gains. Chapter 3 indicates that despite the claims of the balance-of-power and preponderance-of-power schools, neither an even distribution of power nor a preponderance of power is necessarily the most stable distribution. Chapter 4 demonstrates that contrary to the arguments of the power-transition school, power transitions are not the most dangerous phase during a shift in the distribution of power. And, chapter 5 suggests that anarchy does not imply that states generally balance.

Although each of these arguments seems convincing and has been widely accepted, each turns out to be incomplete for one of two basic reasons. In some cases, the argument focuses on one or a few intuitively plausible factors that point in the direction of the purported conclusion. Because these factors do point in the direction of the conclusion, the argument sounds persuasive. But usually there are other equally intuitive factors that point in the opposite direction and would render the argument much less compelling if they were taken into account. However, these opposing factors often remain hidden in ordinary-language arguments because the accounting standards by which this kind of argument is judged frequently do not force opposing factors to the fore. The second related reason that these arguments turn out to be incomplete is that even if the arguments at least initially recognize a fundamental trade-off, they often fail to carry it through the entire analysis.

Formalization helps overcome these problems. Satisfying the accounting standards inherent in specifying a formal model often exposes previously opposing factors and unappreciated trade-offs. (This commonly happens, to the consternation of the modeler who now must contend with a more complicated formulation than was anticipated.) The greater precision of a formal model and the deductive structure underlying it

20 For an excellent example of this evolution of ignorance and the costs and benefits of mathematical models, see Krugman's (1995) discussion of Albert Hirschman's and Gunar Myrdal's rejection of efforts to formalize their ideas and what Krugman calls the "fall and rise of development economics."
also make it easier to trace the implications of this trade-off through the entire analysis and, sometimes, to weigh the relative strength of the competing factors and to predict how these trade-offs are resolved empirically.

**Game-Theoretic Models**

Game theory is a particular kind of mathematical modeling used for studying situations in which a group of actors are strategically interdependent in the sense that each actor's optimal course of action depends on what the other actors will do. These situations are difficult to analyze because deciding what option is best depends on a complex chain of beliefs about beliefs about beliefs about beliefs, and so on. To illustrate these complexities, suppose two actors, \(A\) and \(B\), are in a strategic setting in which each actor's optimal action depends on what the other does. In such circumstances, \(A\) decides what to do on the basis of what it believes \(B\) will do. But what \(B\) does depends on what it believes \(A\) will do. \(A\)'s decision, therefore, is really based on its belief about what \(B\) believes \(A\) will do. But then what \(B\) does is really based on its belief about \(A\)'s belief about \(B\)'s belief about what \(A\) will do, and so on. These chains of beliefs about beliefs make strategic interdependence complicated.

Game-theoretic models help us discipline our thinking about strategic interaction in at least two important ways. The first results from defining a game. Specifying a game tree requires us to describe who the actors are, the order in which they make decisions, what alternatives each actor has to choose from when deciding what to do, and, finally, what each actor knows about what others have done when deciding what to do. These requirements make the assumptions being made about the actors' strategic environment more transparent.

Second, games are generally analyzed in terms of their perfect equilibria. Solving a game for its perfect equilibria disciplines our predictions about how the game will be played, just as defining a game disciplines our thinking about the strategic setting. Equilibrium analysis forces us to look at the situation being modeled from the perspective of each and every actor and to ensure that the prediction makes sense from all of these perspectives.

A perfect equilibrium is a set of strategies—one for each actor—that satisfy two conditions, and meeting these two requirements is what effectively forces us to look at the situation from each actor's position. The first condition is that the set of strategies must be self-reinforcing. That is, no actor can benefit by deviating from its strategy given that the actor believes that all of the other actors are playing according to their strategies. If this condition did not hold, then at least one actor would want to do something other than what he was predicted to do and the prediction as a whole would not make sense. Strategies that satisfy this condition are called *Nash equilibria*.

The second condition is what makes an equilibrium "perfect." This requirement is important because self-reinforcing strategies beg a prior question. A set of strategies is self-reinforcing if no actor can increase its payoff by altering its strategy given that the other actors follow their strategies. But is it reasonable in the first place for an actor to believe that the other actors will play according to the posited strategies? One situation in which it is unreasonable is if the threats and promises implicit in another actor's strategy are inherently incredible. Suppose, for example, that the strategy an actor is presumed to follow relies on a threat which would not be in that actor's own self-interest to carry out if the time came to do so. If other actors know this, then it no longer makes sense for them to assume that the first actor will follow its posited strategy and carry out its threat.

The doctrine of massive retaliation espoused by the Eisenhower administration during the 1950s is a classic example of a set of strategies that are self-reinforcing if believed but are also inherently incredible. In its simplest form, the doctrine said that the United States would launch a massive nuclear attack against the Soviet Union if the Soviet Union precipitated a second Korean War or threatened any other American interest whether that interest be a vital or peripheral American concern. If the Soviet Union believed this threat, then it would not want to challenge the United States. And, as long as this threat deterred the Soviets, it would not have to be carried out and, therefore, making it would have been in the United States' interest. Accordingly, the American strategy of threatening massive retaliation and the Soviet strategy of not challenging the status quo are self-reinforcing. Neither state has any reason to change its strategy if that state believes the...
other will follow its strategy. However, if following through on this strategy would cost the United States more than was at issue in a crisis—as certainly became the case as the United States became increasingly vulnerable to a Soviet nuclear retaliation—then the Soviet Union would have a good reason to doubt that the United States actually would carry out the threat and the doctrine of massive retaliation would be incredible.  

Insisting that a set of self-reinforcing strategies also be perfect helps resolve this issue formally. Perfection requires that following through on the threats and promises implicit in each actor’s strategy be in that actor’s self-interest. Thus, no actor has any reason to doubt that any other actor will not play according to its posited strategy.

Assessing Models

Three considerations need to be kept in mind when evaluating game-theoretic models and, especially, those developed below. The first relates to the art of modeling. The accounting mechanism embodied in game theory is very limited and can easily be overwhelmed by trying to incorporate too many factors in a single formalization. Adding more and more elements to a model may bring some advantages. It may appear to make that model more general and better able to capture the complexities of an empirical situation. These benefits, however, must often be weighed against some significant costs. Incorporating many factors in a game can readily render the model utterly intractable. Including too much can make a model too complicated to analyze. Even if the model can be analyzed formally, it may still be difficult or impossible to understand the role of one or two factors if too many other factors have been included. Although less general and less representative, a simpler model that has been well designed to focus on one or two factors may actually prove to be more useful and insightful. The art of modeling is finding a formulation that strikes an acceptable balance between these costs and benefits. The need to strike this balance suggests that the most important criticism of a model is not that it is simple or leaves much out. Rather, the important issue is whether what a model leaves out seems likely to affect the conclusions that are being derived from it, and, if so, how these factors might be incorporated in an interesting and tractable way.

This judgment should also be made comparatively. A model may be too simple in some absolute sense and one may hope to relax some restrictive assumptions as the modeling enterprise continues. But how does the present model compare to existing alternatives? Formal models in international relations theory do surprisingly well by this comparative standard. Many ordinary-language analyses in international relations theory are described in the context of complicated historical cases. But when one strips away this descriptive richness to examine the underlying causal structure, that structure often turns out to be very simple. The more threatening a state is, the harder other states will try to counter or balance against it. A preponderance of power is more peaceful because a very weak state is unlikely to prevail if it resists the demands of a much stronger state. An offensive advantage raises the payoff to attacking relative to being attacked and this makes war more likely. If waiting while others fight can rapidly lead to an adverse shift in the distribution of power, states will be less likely to wait and will appear to be chain-ganged together. Formal models look very stark in comparison to the descriptive and historical detail of many ordinary-language analyses. But this apparent contrast between “rigor and richness” is often a matter of presentational style and not causal complexity. When one looks at the underlying causal arguments, formal models in international relations theory frequently are at least as rich causally as ordinary-language analyses. Indeed, the accounting mechanism inherent in the formalization may make it possible to see more complex causal relations.

Second, formal models are often criticized because so many assumptions have to be made in order to specify the model, and the conclusions usually depend on which assumptions are made. This may make a formal analysis appear to be much less robust than an ordinary-language analysis. This criticism is, however, a bit like shooting the messenger because one does not like the message. If assumptions are important, they are important whether we recognize it or not. Formalization does not make these assumptions important; it only helps us see that they are. And, failing to take important assumptions and the conditionality inherent in them into account may lead to explanations and claims that are too broad and do not hold up when the modeling dialogue moves back to the more empirical realm to consider new case studies or statistical analyses.

24 Although he did not use the language of game theory, this is the basic logic behind Kaufmann’s (1956) criticism of massive retaliation. See Powell (1990, 12–32) for a more extensive discussion of this example.
The third consideration centers on the broader contribution that models and, especially, game-theoretic models can make and on a criticism of this contribution. As discussed above, formal analyses at their best do much more than show that existing arguments fail to go through because they are wrong or incomplete. The best models suggest new insights and new ways of looking at things. These insights are most useful when they go beyond the models that produced them and help explain or illuminate a wide range of substantive problems. Once seen, these ideas often seem quite intuitive. Indeed, they almost have to be intuitively clear if they are to be widely applied. But, ironically, the fact that the insights are intuitively clear—at least in retrospect—leaves them vulnerable to the charge that the models were unimportant and not really essential to generating the insights in the first place. One could have come up with these intuitive ideas without laboring through a formal analysis.

This criticism is correct in principle but seems to be wrong as a matter of fact. In some instances the accounting mechanism embodied in a formal model provides a simpler way of working with a complicated set of issues. This is precisely what models have to offer and why people use them. But any formal argument can be translated into ordinary language. One can translate an equation into English. Thus, any conclusion derived from a formal model can in principle be derived from an ordinary-language analysis. But what is possible in principle frequently does not occur in practice. As much of the recent formal work in international relations theory shows, formal models have often proved to be an important source of new insights. Perhaps these insights could have originated in ordinary-language analyses, but the fact is that they did not.

Some Templates

The following chapters are part of modeling dialogue. They use a series of game-theoretic formalizations to try to discipline and deepen our understanding of international politics by examining three ways that states can respond to threats. The models are very spare and, indeed, almost certainly too spare to explain any particular outcome in any degree of specificity. Instead, the models are intended to provide a kind of template that helps us organize our thinking about specific problems and more general issues. This template offers a framework and point of departure for the analysis. It identifies critical assumptions and explores some basic issues. But just as the present models qualify some of the existing arguments in international relations theory, future work will undoubtedly qualify some—if not all—of the conclusions derived from the present models by showing how different assumptions may lead to different conclusions. This is the way that the modeling enterprise works, and the rest of this book tries to lay some of the foundation for that future work and for the next round of the modeling dialogue.