
In the Shadow of Power

States and Strategies in International Politics

ROBERT POWELL

1999

PRINCETON UNIVERSITY PRESS

PRINCETON, NEW JERSEY

Copyright ©1999 by Princeton University Press
Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
Chichester, West Sussex

All Rights Reserved

Library of Congress Cataloging-in-Publication Data

Powell, Robert, 1956-

In the shadow of power : states and strategies in international
politics / Robert Powell.

p. cm.

Includes bibliographical references and index.

ISBN 0-691-00456-0 (cl. : alk. paper). — ISBN 0-691-00457-9 (pbk. : alk. paper)

1. International relations. I. Title.

JZ1242.P68 1999

327.1'01—dc21

99-12207

This book has been composed in Times Roman and Times Roman Bold

The paper used in this publication meets the minimum requirements of
ANSI/NISO Z39.48-1992 (R1997) (*Permanence of Paper*)

<http://pup.princeton.edu>

Printed in the United States of America

1 3 5 7 9 10 8 6 4 2

1 3 5 7 9 10 8 6 4 2

(Pbk.)

Contents

Preface	vii
Chapter 1 – States and Strategies	3
Chapter 2 – Guns, Butter, and Internal Balancing in the Shadow of Power	40
Chapter 3 – Bargaining in the Shadow of Power	82
Chapter 4 – Bargaining in the Shadow of Shifting Power	115
Chapter 5 – Alignment Decisions in the Shadow of Power	149
Chapter 6 – Conclusion	197
Appendix 1 – Game Trees, Strategies, and Equilibria	225
Appendix 2 – The Formalities of the Guns-versus-Butter Model	233
Appendix 3 – The Formalities of Bargaining in the Shadow of Power	243
Appendix 4 – The Formalities of Bargaining in the Shadow of Shifting Power	272
Appendix 5 – The Formalities of the Alignment Model	283
References	291
Index	305

States and Strategies

How does the shadow of power affect state behavior? How, for example, does the distribution of power affect the likelihood of war? Is war less likely if there is an even distribution of power or if one state preponderates? Or, is there any relation at all between the distribution of power and the risk of war?

How do states cope with shifts in the distribution of power among them? Thucydides ascribed the cause of the Peloponnesian War to the rise of Athenian power: "What made war inevitable was the growth of Athenian power and the fear which this inspired in Sparta." But the profound shift in power that took place between Britain and the United States in the century after the War of 1812 did not lead to war even though it was at times fraught with conflict and tension. Why do some shifts lead to war and others not? When is a declining state tempted to launch a preventive war against a rising state, and when does it heed Bismarck's counsel that "preventive war is like suicide from fear of death" by trying to accommodate a rising state?

If one state threatens another, how do other states react? Do they tend to balance against the threat by aligning with the threatened state as republican France and czarist Russia did against Germany in 1894? Or do states tend to bandwagon by joining in with the state making the threat as Prussia, Austria, Russia, and Spain did at various times during the Napoleonic Wars? Or do states try to avoid conflict by not taking sides as the United States did for the first three years of the First World War?

How do changes in military technologies affect the prospects for peace? Is war more likely if technology gives the offense an advantage over the defense, as was believed to be the case in the summer of 1914? Or do the states, recognizing the greater technological danger, take steps to offset it as states tried to do before the First World War by concentrating their troops farther back behind the frontier? And, if it such steps are unavailable, do states compensate for the greater

technological danger by moderating their behavior so that the overall risk of war remains largely the same?¹

These were important questions before the end of the cold war and they remain so afterward. China's rapid economic growth since Mao's death has begun to raise security concerns about East Asia. The demise of the Soviet Union has left a power vacuum in resource-rich Central Asia and has sparked talk of another round of the "Great Game." And, some ponder the effects of the possible reemergence of Russia as a Great Power.

Much of international relations theory seeks to understand how states interact in the shadow of power by trying to abstract away from many of the details of international politics and foreign policy. These analyses focus instead on attempting to understand the strategic logic of a simple, stylized model of the international system. This stylization strives to distill the essence of some of the strategic problems states actually face—or at least important aspects of them—into a simpler analytic setting where it is easier to see how states deal with these problems, how these efforts interact, and what the outcomes of these interactions are.

In this stylization, a small number of states interact in a Hobbesian state of nature. Each state is assumed to be a rational, unitary actor and to have control over a limited amount of resources. These resources can be used in one of two general ways. A state can allocate resources toward the satisfaction of its immediate ends or toward the means of military power. Because these states exist in a condition of anarchy or a Hobbesian state of war, there is no supranational Leviathan to impose order and to protect the states from each other. There is nothing to stop one state from trying to further its interests by using its military means against another state if the former believes that doing so is in its best interest. However, a state's ability to achieve its ends through the use of military force against another depends in part on the relative amounts of resources the states have devoted to the military. The more a state allocates to the military, the stronger it becomes *ceteris paribus*

¹See, of course, *The Peloponnesian War* for Thucydides' account and see Kagan (1969, 1974, 1981, 1987) for a modern treatment. Allen (1955), Bourne (1967), and Campbell (1974) provide an overview of the relations between Britain and the United States. Levy (1987, 103) quotes Bismarck's view on preventive war. Kennan (1979, 1984) discusses the Franco-Russian relations and A. Taylor (1963, 325–44) offers a brief summary. Schroeder (1994a, 1994b) and Schweller (1994) examine balancing and bandwagoning. Jack Snyder (1984a, 113) notes France's efforts to decrease its vulnerability by locating its railhead disembarkation points far enough back from the frontier. Nalebuff (1986) studies the possibility that states moderate their behavior when the technological risks are higher.

and the more likely it is to prevail in any conflict. Thus, allocations to the military, which are assumed to have no intrinsic value and bring no direct gain to a state, may bring important indirect gains by providing a state with the means of protecting the resources it already has or of taking resources from others.

Variants of this stylization and efforts to understand it can be found in Thucydides' account of the Peloponnesian War; in E. H. Carr's ([1939], 1954) study of the twenty-years' crisis between the two world wars; in Quincy Wright's ([1942], 1965) compendium on war; in Martin Wight's (1946) examination of power politics; in Hans Morgenthau's early attempt to "present a theory of international politics" ([1948], 1967, 3); in Edward Gullick's (1955) endeavor to systematize balance-of-power theory; in Morton Kaplan's (1957) application of "systems theory" to international politics; in John Herz's (1950) and Herbert Butterfield's (1950) early explications of the security dilemma; in Kenneth Waltz's (1959) third image of the causes of war; in Arnold Wolfers's (1962) theorizing about discord and collaboration; in Raymond Aron's (1966) "determination to offer a general theory, starting from the specific features of international relations" (Hoffmann 1987, 53); in Hedley Bull's (1977) striving to understand international order and an anarchical society; in Robert Keohane and Joseph Nye's (1977) work on power and interdependence; in Robert Jervis's (1978) elaboration of the security dilemma and its relation to the offense-defense balance; in Waltz's (1979) attempt to construct a structural theory of international politics; in Bruce Bueno de Mesquita's (1981) start on formalizing the expected-value calculations leading to war; in Robert Gilpin's (1981) explanation of hegemonic war; as a point of departure for Keohane's (1984) efforts to understand international cooperation and the role of institutions; in Robert Axelrod's (1984) study of cooperation; and, albeit often less explicitly so, in a myriad of other analyses. Indeed, this stylization has now become part of what Thomas Kuhn (1970) would call the "disciplinary matrix" of international relations theory.

In this stylized environment, a state can respond to threats to its security in at least three general ways. It can reallocate the resources already under its direct control in what Waltz calls "internal balancing" (1979, 168). A state can try to resolve conflicts and diffuse threats through bargaining and compromise. Or, it can try to draw on the resources of others by allying with them.

This book tries to advance our understanding of the strategic logic of this stylized environment and of international politics more broadly

by examining each of these three responses. More fundamentally, the analysis tries to illuminate some of the basic mechanisms through which states pursue their ends and conflicts of interest play themselves out. These mechanisms are too general and too spare to explain particular outcomes in any degree of specificity. These mechanisms instead provide a kind of template for thinking about the ways that states—or, more broadly, actors—interact in the shadow of power.

This examination suggests that many widely made arguments about the ways that states interact in this environment and about the outcomes of this interaction are at best incomplete and need to be qualified: Neither a balance nor a preponderance of power is necessarily more peaceful. There is no general tendency to balance. Anarchy does not imply that states are concerned about relative gains and that these concerns in turn impede cooperation.

The following chapters indicate instead that the relation between the likelihood of peace and the distribution of power depends on the distribution of benefits associated with the existing international order. War is least likely when the international distribution of benefits reflects the underlying distribution of power. The greater the disparity between the distribution of power and the distribution of benefits, the more likely war. Whether states balance, bandwagon, or stand aside while others fight depends in a complicated way on many different factors, which include the cost of fighting, the extent to which military forces cumulate when combined together in an alliance, and the relative aggressiveness of potential coalition partners.

The remainder of this chapter serves three ends. First, the analysis of the three responses to threats emphasizes the importance of commitment issues, the effects of asymmetric information, and what will be called the “technology of coercion.”² The next section introduces these ideas. The subsequent section then provides an overview of each of the following chapters. These chapters use a series of game-theoretic models to examine the strategic problems and trade-offs states face when responding to threats.³ This modeling-based approach raises a methodological question, which the final section of this chapter addresses: What is the role and usefulness of formal models in the study of international politics?

²See Fearon (1995b) for a discussion of the importance of commitment issues and informational asymmetries in theories about the causes of war.

³As elaborated more fully below, the general discussion and analysis of these models does not presuppose any previous knowledge of game theory. The technical analysis of the models is confined to the appendixes.

Commitment Issues, Informational Problems, and the Technology of Coercion

Commitment issues, informational asymmetries, and the technology of coercion play key roles in defining the strategic problems states face, and this section introduces these terms. A *commitment problem* exists among a group of actors if they could make themselves better-off by being able to commit themselves to following a particular course of action but are unable to do so. This situation occurs when, loosely speaking, individually sensible decisions lead to outcomes that are collectively inferior. More precisely, commitment problems exist if each actor’s acting in its own self-interest leads to an outcome that leaves everyone worse-off.

The prisoner’s dilemma is a classic example of a commitment problem, and it has been widely used in international relations to illustrate problems as diverse as the Concert of Europe (Jervis 1985), the failure of cooperation in the July 1914 crisis (Van Evera 1985), arms races (Downs and Rocke 1990), trade agreements and wars (Conybeare 1985, 1987; Yarbrough and Yarbrough 1992), international banking (Lipson 1985), alliance politics (G. Snyder 1984), and the effect of alliances on trade (Gowa 1989, 1994; Gowa and Mansfield 1993). In the prisoner’s dilemma, two players, say *I* and *II*, must choose between the two alternatives of “cooperating” and “defecting.” Suppose for example that the actors are states and that each state can decide to build up its arms or not build. These choices are illustrated in figure 1.1 where *I*’s choices are denoted by rows and *II*’s choices are represented by columns. If, for instance, *I* decides to build and *II* does not, then the outcome would be in the upper-right cell of the matrix. The four cells of the matrix show all of the possible combinations of the states’ decisions and, implicitly, the outcomes associated with these decisions.

The numbers in the cells reflect the way the states rank the possible outcomes. In a prisoner’s dilemma, an actor’s most preferred outcome is for it to defect while the other actor cooperates. In the arms-race context, this means that each state’s most preferred outcome is for it to build while the other does not because this is assumed to lead to a favorable shift in the distribution of power. The next best outcome is for neither state to build, as this leaves the distribution of power unchanged and avoids the added economic burden of spending more on the military. The third-best outcome is for both states to build. This, too, leaves the distribution of power unchanged but now entails spending more on arms. And, the worst outcome for a state is not to build when the other does, as

		State II	
		Build	Not Build
State I	Build	2,2	4,1
	Not Build	1,4	3,3

Figure 1.1 The decision to build armaments as a prisoner's dilemma

this leads to an unfavorable shift in the distribution of power. Using the numbers 1 through 4 to depict this ranking and letting larger numbers correspond to more favorable outcomes then gives the entries in the cells in figure 1.1 where the first number in a pair represents *I*'s ranking.

To see that the states face a commitment problem in the prisoner's dilemma, consider what they will do. If the actors are unable to commit themselves to not building, then each builds. This follows from the observation that if *II* builds (i.e., chooses the left column in figure 1.1), *I* obtains a payoff of two by building and a payoff of one by not building. *I*, therefore, prefers to build if it expects *II* to build, as this prevents an unfavorable shift in the distribution of power.

In fact, *I* also does better by building even if *II* does not build (i.e., if *II* plays the right column). If *I* builds while *II* does not, the distribution of power shifts in *I*'s favor. By contrast, *I* does save money if it does not build, but it foregoes the opportunity to shift the distribution of power in its favor. The assumptions that we have made about state preferences in this simple example then imply that *I* prefers the former outcome to the latter.⁴ In numbers, *I* obtains 4 if it builds and 3 if it does not. Thus, *I* prefers to build even if *II* does not build.

⁴What to assume about state goals and preferences has recently been the subject of intense debate in international relations theory, and will be discussed more fully below in chapter 2.

In sum, *I* does better by building regardless of what *II* does. A similar argument shows that *II* also prefers to build regardless of what it expects *I* to do. Consequently, each state builds and receives a payoff of two as long as the states are unable to commit themselves to refraining from building. However, each state would have done better if both had been able to deny themselves the option of building as each would have received a payoff of three. The fundamental problem facing actors caught in a prisoner's dilemma arises because they are unable to commit themselves to a course of action that would benefit both of them.

A commitment problem also lies at the center of the security dilemma (Butterfield 1950; Herz 1950; Jervis 1978; Glaser 1992, 1997). A security dilemma exists when one state's efforts to increase its security also has the—perhaps unintended—effect of reducing another state's security. In an effort to restore its initial level of security, the second state may then take steps that diminish the effectiveness of the first state's efforts to enhance its security. This, in turn, prompts the first state to take additional steps and so on. In a security dilemma, one state's efforts to increase its security may trigger a spiral of hostility leading to war. In this situation, both states would be better-off if the first state could somehow commit itself to refraining from using its greater military capability offensively. If such a commitment could be made, then the first state's initial efforts to increase its security would not spark a spiral by threatening the second state.

The second feature that often characterizes the strategic problems states face is *asymmetric information*. Information is asymmetric or incomplete when different actors know or believe different things about a situation. These asymmetries typically take the form of uncertainty about states' goals (or, more formally, their preferences) or capabilities. For example, each state may know how much it values prevailing in a crisis but be unsure of how much another state values prevailing. Information is asymmetric in this case because each state knows its own payoff to prevailing but not that of the other state. States may also have different beliefs about their capabilities and therefore about their chances of prevailing. When each state believes that it is well prepared and, lacking complete information about the other side, also believes that the other side is comparatively less well prepared, then both states will overestimate their chances of prevailing. Conversely, each state may believe itself to be unprepared for war while fearing that the other is. In this case, the informational asymmetry leads both

states to underestimate their chances of success and thereby facilitates a peaceful resolution of a dispute.⁵

Informational asymmetries about states' goals abound in international politics. Uncertainty about the scope of Hitler's ultimate ends was a crucial aspect of the strategic problem facing Britain in the 1930s. Britain was generally willing to revise the Versailles Treaty in Germany's favor (Gilbert 1966) but unwilling to let Germany dominate the continent. However, Britain was unsure whether Hitler's aims were limited or not. As Alexander Cadogan, the new permanent under-secretary in the British Foreign Office put it shortly after Germany annexed Austria: "I am quite prepared to believe that the incorporation in the Reich of Austrian and Sudetendeutsch may only be the first step in a German expansion eastwards. But I do not submit that this is necessarily so, and that we should not rush to hasty conclusions."⁶ This informational asymmetry led Britain to make a series of concessions to Germany throughout the 1930s in the hope that Hitler's ends were limited and that these concessions would appease him. But each concession led to a new demand and, ultimately, war when Britain became convinced that Hitler's aims were unlimited.

Asymmetric information about goals also played a critical role in the Agadir crisis, which raised the specter of war between Britain and Germany in the summer of 1911 and subsequently led to an intensification of the arms race (Barlow 1971; Barraclough 1982; Kennedy 1980; Stevenson 1996; A. Taylor 1963; S. Williamson 1969). In May of that year, France tried to enhance its position in Morocco by occupying Fez, an occupation which, if prolonged, would violate the Algeçiras Act. Earlier that spring the French ambassador to Berlin had warned his government that violating that act would lead Germany to demand a Moroccan port in compensation, and the crisis grew out of the bargaining over Germany's compensation.

German Foreign Minister Alfred von Kiderlen-Wächter exacerbated tensions when he tried to demonstrate Germany's resolve by order-

⁵Much work in international relations has been done on the effects of misperception (e.g., Jervis 1977; Levy 1983; Stein 1982), and the concepts of misperception and asymmetric information are related. Both approaches posit actors that may have different beliefs about a situation. But there is also an important and complementary difference in emphasis. The former usually focuses on psychological processes and cognitive limitations, which may lead to a distorted view of a situation. The latter, by contrast, focuses on the uncertainty that arises simply because actors know different things about a situation and not because they process information differently.

⁶Quoted in R. A. C. Parker (1993, 135).

ing the gunboat *Panther* to Agadir. But what drove the crisis forward was "Kiderlen's failure to spell out his objectives to the other Powers" and their lack of information about those objectives (Stevenson 1996, 183–92). The problem was that if

(as the German evidence suggests) he [Kiderlen] wished to abandon Morocco and get the best possible Central African compensation, all three capitals could perhaps be satisfied. If he wanted a political presence in the sultanate he might be able to drive London and Paris apart. But if he intended to inflict on France a humiliation that would jeopardize its Great-Power status, the entente would come back into play. (Stevenson 1996, 184)

This uncertainty led Britain to interpret Kiderlen's disproportionately large demand for all of the French Congo as an attempt to humiliate France and ultimately as a threat to the entente between Britain and France. As a result, British Foreign Secretary Grey began to take a firmer line with Germany; Lloyd George, with Grey's approval, publicly warned in his Mansion House speech that Britain's interests had to be taken into account; and a war scare between Britain and Germany followed.

Asymmetric information about capabilities also pervades international politics. When Russia declared war on Turkey in April 1887, both Britain and Russia were confident of a Russian victory. "In June it looked as though the Russians would conquer all of Turkey-in-Europe within a month; then they ran against the *hitherto unknown* fortress of Plevna, barring the road further south, and failed to take it until 11 December" (A. Taylor 1963, 245, emphasis added). By the time Russia reached Constantinople in January, its army was exhausted and the Ottoman Empire had been given "another forty years of life" (A. Taylor 1963, 245). Two decades later the Russo-Japanese war resulted in part from Russia's underestimation of Japan's military strength and a general unwillingness to make concessions that arose from it (Fearon 1995b, 398–400; Lebow 1981, 244–46; Nish 1985, 240–41; Richardson 1994, 292; White 1964, 95–146). "A widespread view in Port Arthur was that the Japanese army would not dare to attack because of its poor calibre" (Nish 1985, 241).

Britain's decision to fight on alone in the summer of 1940 also seems to have been based at least in part on asymmetric information about Germany's capabilities. British leaders mistakenly "believed that the German war machine was approaching maximum efficiency" and was therefore vulnerable to economic pressure and strategic bombing

(Reynolds 1985, 160).⁷ More generally, asymmetric information underlies Geoffrey Blainey's discussion of the causes of war. He argues that uncertainty about the distribution of power is a major cause of war: "Wars usually result when the fighting states *disagree* about their relative strength" (1973, 122) and are overly optimistic about their prospects for success.⁸

Informational asymmetries create strategic problems when two conditions hold. The first is that the missing information must matter. That is, at least one actor would alter its behavior if it knew what another actor knows. Uncertainty does not matter if resolving that uncertainty would not affect the actors' actions. Second, when one actor knows something another does not, the former must have some incentive to misrepresent or, more bluntly, lie about what it knows (Fearon 1995b, 395–401). Otherwise, each actor could simply ask the other actors to disclose what they know.

A simple bargaining model, which will be studied at length in chapter 3, illustrates these two conditions. Suppose that a satisfied state and a dissatisfied state are bargaining about revising the territorial status quo. The situation is depicted in figure 1.2. The line between the satisfied and dissatisfied states' capitals represents the territory to be divided. The satisfied state currently controls the territory to the right of the status quo q , and the dissatisfied state controls the territory to the left of q .

Assume further that the satisfied state knows that the other state is dissatisfied with the present territorial division and will go to war to change it unless a compromise is found. But, the satisfied state is unsure as to how much it has to concede to the other state to satisfy its minimal demands and thereby avert a war. To keep the example simple, assume that there are two possibilities: the dissatisfied state's demands are either moderate or extensive. If the dissatisfied state's demands are moderate, then its demands would be satisfied if it controlled the territory to the left of m . That is, while this state prefers to control more territory than less, it is unwilling to use force to expand its border beyond m . If, by contrast, the dissatisfied state's demands are extensive, then it has to control the territory to the left of e before it becomes unwilling to use force to acquire more territory. Finally, suppose that the satisfied state

⁷For other discussions of British intelligence, see Hinsley et al. (1979–90) and Wark (1985).

⁸Morrow (1989) and Bueno de Mesquita, Morrow, and Zorick (1997) analyze this problem formally.

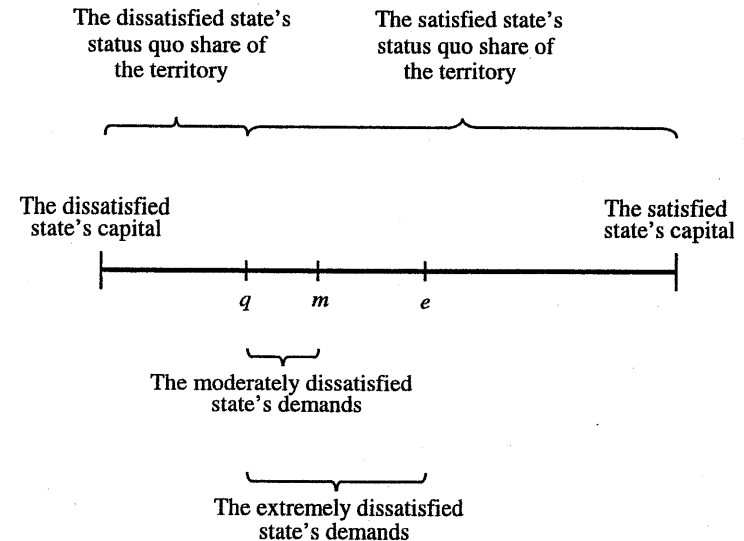


Figure 1.2 Territorial bargaining

is willing to withdraw to m or e to avoid war but prefers holding the line at m if that would be enough to appease the other state.

In this situation, the presence of asymmetric information matters. If the satisfied state knew what the dissatisfied state knows, i.e., whether the dissatisfied state's demands are moderate or extensive, then this knowledge would affect the satisfied state's behavior. If, for example, the satisfied state knew that its adversary's demands were moderate, then the satisfied state would agree to withdraw to m but not to e . Moving back to m would be enough to meet the dissatisfied state's minimal demands and eliminate the risk of war. If, by contrast, the satisfied state knew that the dissatisfied state's demands were extensive, the satisfied state would withdraw to e to remove the danger of war.

The fact that the satisfied state would change its behavior if it knew what the dissatisfied state knows does not in and of itself create a strategic problem. If the dissatisfied state has no incentive to misrepresent the scope of its demands, then the satisfied state could simply ask its adversary whether its demands were moderate or extensive. If the answer was moderate, then the satisfied state would pull back to m ; if the dissatisfied state declared that its demands were extensive, the satisfied state would retreat to e . If, however, the satisfied state is going to react in this way

to whatever the dissatisfied state says, then this gives the latter an incentive to claim that its demands are extensive even if they are not. Put more bluntly, if the dissatisfied state's demands actually are moderate but the satisfied state believes whatever the dissatisfied state says, then the latter can gain by bluffing. The incentive for an actor to misrepresent what it knows is also a critical element in transforming an informational asymmetry into a strategic problem.

In sum, the satisfied state's behavior would differ if it knew whether its adversary's demands were moderate or extensive. Moreover, the dissatisfied state also has an incentive to misrepresent its demands if they are moderate, and therefore the satisfied state cannot simply rely on what the other says its demands are. In these circumstances, the satisfied state's uncertainty about the dissatisfied state's demands creates a trade-off. The satisfied state could be sure of obtaining a peaceful settlement by offering extensive concessions that would satisfy the other state regardless of its actual demands. Alternatively, the satisfied state could offer only moderate concessions by drawing the line at m . The advantage of this is that if the dissatisfied state's demands really are moderate, it would accept this new territorial division and the satisfied state would not have to give up as much. The obvious disadvantage to making moderate concessions is that if the dissatisfied state's demands actually are extensive, then standing firm at m means war. The satisfied state, therefore, faces a trade-off. It can increase the chances of a settlement by offering more, but, because it has offered more, the agreement will be less attractive. This kind of trade-off is typical of situations in which there is asymmetric information and will play a crucial part in the analysis in the following chapters.

Along with commitment issues and informational asymmetries, the *technology of coercion* helps define the strategic arena in which states interact and the problems they face there. The technology of coercion is, roughly speaking, the analogue of a production function in microeconomics. A production function describes the technological relation between inputs and outputs; it specifies, for example, how much more labor it takes to produce one more unit of output. *The technology of coercion describes the relation between what an actor does and how those actions exert coercive pressure.* The technology of coercion thus defines at least implicitly what it means to be powerful in a particular situation. The greater an actor's coercive capabilities, the more powerful it is.

Changes in this technology can have important effects on the outcome of state interactions. Two examples illustrate the point.⁹ First, the balance between offense and defense has long been thought to have a significant effect on states' interaction. The offense has the advantage when "it is easier to destroy the other's army and take its territory than it is to defend one's own. When the defense has the advantage, it is easier to protect and to hold than it is to move forward, destroy, and take" (Jervis 1978, 187). Jervis (1978), Quester (1977), Van Evera (1998), and others argue that offensive advantages make war more likely and the international system less peaceful. Christensen and Snyder (1990) also suggest that the offense-defense balance affects the alliance behavior of states. When states believe that the offense has the advantage, as in the summer of 1914, alliances are tighter and states are more likely to be dragged into a war by an adventurous ally. By contrast, alliances are looser and the members of an alliance work harder to ensure that the other members of the alliance bear most of the cost of resisting an opposing alliance when the defense is perceived to have the advantage, as in the 1930s.

The second example of the effects of a change in technology is the nuclear revolution. The advent of secure, second-strike nuclear forces created a condition of mutually assured destruction (MAD), in which both the United States and the Soviet Union could launch a massive, society-destroying second strike after absorbing a first strike from the other. This technological development made it impossible for a state to defend itself and in so doing changed the way that states exert coercive pressure on each other. Before the nuclear revolution, the use or threatened use of force during a confrontation was a contest of military strength in which each state tried to defend itself while destroying the other's military capabilities. By rendering defense impossible, the nuclear revolution transformed the way that coercive pressure is brought to bear. Because each state remains vulnerable to a society-destroying attack from the other regardless of the verdict on the battlefield, the means through which coercive pressure can be applied changes from a contest of military strength into a contest of resolve (Schelling 1966). Each state now tries to influence the other during a crisis by trying to hold on longer in the face of a growing risk that events will go out of

⁹McNeill (1982) offers a broad historical survey of some of the effects of military technology.

control.¹⁰ One consequence of this change in the technology of coercion, it is argued, is that it makes major war much less likely (Jervis 1984; Waltz 1990).

The two previous examples, like most of this book, focus on the use or threatened use of military force to exert coercive pressure and on the effects that changes in that technology have on these efforts. Nevertheless, the broader term “technology of coercion” is used instead of the perhaps more direct, but also more limited, term “military technology.” The former generalizes the latter by describing how actors can exert coercive pressure on other actors in settings where the use of military power is not a significant concern. This broader notion, along with commitment issues and informational asymmetries, may be useful in characterizing political situations in international relations and politics more generally where coercion and the exercise of power are very much at issue but in which the use of military force is not.

Three Responses to Threat

Once we begin to think about international politics in terms of commitment issues, informational problems, and the technology of coercion, what patterns should we expect to see? The following chapters address this question by examining three ways that states can respond to threats in the context of the simple stylization of the international system described above. Chapter 2 focuses on internal balancing and the guns-versus-butter problem underlying it. When a state tries to further its ends and maintain its security by relying on its own limited resources, what determines how much of these resources it devotes to the means of military power? How do the states’ decisions interact, and when do these interactions sustain a peaceful outcome and when do they break down in war?

A commitment issue lies at the center of the guns-versus-butter problem. Because states are unable to commit themselves to refraining from using force against each other, they face a fundamental trade-off between the resources they devote to satisfying their immediate ends and those they allocate to the means of military power. The more a state allocates to the military, the less it can devote to satisfying its immediate wants. However, allocating more to the military today makes that

state stronger tomorrow. Being stronger, in turn, makes it easier for that state to defend the resources it has and, if it wishes, to capture additional resources. Conversely, the more a state devotes to satisfying its immediate wants, the less it can allocate to the military and the weaker it will be in the future. Should another state then decide to attack, this military weakness means a higher probability of defeat and, in the event of defeat, fewer future resources to devote to its immediate ends. The specter of war, therefore, gives rise to a trade-off between the present and the future. The more resources a state devotes to satisfying its current wants, the fewer resources it can expect to have in the future.

This trade-off would disappear if the states could commit themselves to refraining from using military force against each other. If such commitments were possible, no resources would have to be spent on deterring or defending against an attack. Instead, each state could devote those resources toward satisfying its ends, and this reallocation would make each state better-off. The states’ inability to deny themselves the option of using force makes all of them worse-off. This is a commitment problem.

The analysis of the guns-versus-butter trade-off highlights three issues. First, the analysis clarifies the circumstances in which internal balancing can maintain the peace and why it may break down in war. The second issue is the effect of changes in the technology of coercion on the guns-versus-butter trade-off. The analysis helps explain how changes in the offense-defense balance and in the cost of fighting affect the states’ military allocations. As will be seen, shifts in favor of the offense lead to larger allocations whereas higher costs lead to smaller allocations.

The third issue that the guns-versus-butter trade-off highlights is the effect of relative-gains concerns on the prospects for international cooperation. Waltz argues that when states are considering a cooperative agreement in a system in which they are unable to commit themselves to refraining from using force against each other, then those states “are compelled to ask not ‘Will both of us gain?’ but ‘Who will gain more?’ . . . Even the prospect of large absolute gains for both parties does not illicit their cooperation so long as each fears how the other will use its increased capabilities” (1979, 105).

The claim that states must be concerned with their relative gains has led to a heated debate about the effects of these relative-gains concerns on the ability of states to realize mutual gains through international

¹⁰For discussions of the nuclear revolution, see Brodie (1946, 1959), Herz (1959), Schelling (1966), Jervis (1984), and Powell (1990).

cooperation.¹¹ The analysis of the guns-versus-butter problem suggests that the effects of the relative-gains concerns have been vastly overstated and generally do not impede cooperation.

Chapters 3 and 4 focus on states' efforts to respond to threats through bargaining and compromise. To foreshadow the discussion, suppose that a satisfied and a dissatisfied state are bargaining about revising the territorial status quo. If the states had complete information about each other, then the satisfied state would know the other state's payoffs. In particular, the satisfied state would know how much it has to offer the other state in order to satisfy its adversary's minimal demands and thereby remove the threat of war. The satisfied state thus faces a simple, if stark, choice. It can either make the needed concessions or fight.

Asymmetric information complicates the situation. If the satisfied state is unsure of what it takes to meet the other's minimal demands, the satisfied state faces a trade-off. The more it offers, the greater the chance of appeasing the other state and the smaller the probability of war. But the more the threatened state offers, the less it will have in the event that its offer is accepted.

Chapter 3 examines how changes in the distribution of power and the technology of coercion affect this trade-off and the likelihood of war. The balance-of-power school argues that the probability of war is smallest if there is an even distribution of power among the states (Claude 1962; Mearsheimer 1990; Morgenthau 1967; Wolfers 1962; Wright 1965), whereas the preponderance-of-power school claims that the risk of war is smallest if one state has a preponderance of power (Blainey 1973; Organski 1968; Organski and Kugler 1980). The analysis in chapter 3 qualifies these claims and shows that neither a balance nor a preponderance is necessarily more peaceful.

War is least likely when the international distribution of benefits reflects the underlying distribution of power. The larger the disparity between these distributions, the more likely war. Consequently, the risk of war is low *if* there is an even distribution of power and *if* there is also an even distribution of benefits but not *if* the distribution of benefits favors a particular state. Similarly, the risk of war is low *if* one state has a preponderance of power and *if* that state also receives a preponderance of the benefits but not *if* there is an even distribution of benefits.

¹¹This debate will be reviewed in more detail in chapter 2. For a sampling of it, see Grieco (1988a, 1993); Grieco, Powell, and Snidal (1993); Keohane (1993); Powell (1991, 1993, 1994); and Waltz (1979).

Changes in the technology of coercion often have simple overall effects on the likelihood of war, but the simplicity of the overall effect masks a complicated interaction of opposing factors. Consider, for example, the consequences of a shift in the technology of coercion, which makes fighting more costly. Chapter 3 shows that this shift affects the bargaining between the satisfied and dissatisfied states in three ways. First, the dissatisfied state is less likely to use force because fighting is more costly, and this tends to make war less likely. But, the satisfied state, knowing that its adversary is less likely to use force, tends to offer smaller concessions, and this factor makes war more likely. However, fighting is also more costly for the satisfied state and this tends to produce larger concessions, which makes war less likely. As we will see, the net effect of an increase in the cost of fighting is a lower risk of war. But when we look closely at the mechanism through which states exert coercive pressure on each other, it is not at all obvious that things should work out in this simple way.

A shift in the offense-defense balance also influences the bargaining in opposing ways. Larger offensive advantages incline the satisfied state to make larger concessions. But these advantages also make the dissatisfied state less likely to accept any specific offer because the larger offensive advantages raise the expected payoff to attacking. The overall effect of larger offensive advantages is to make war more likely.

Thucydides attributed the Peloponnesian War to the rise in Athenian power, and chapter 4 generalizes the model developed in chapter 3 to examine how two states cope with a shift in the distribution of power between them. Suppose that the distributions of power and benefits initially favor one of the states. As long as the distribution of power remains constant, then the distribution of benefits will continue to reflect the underlying distribution of power and there will be peace. But what happens if the distribution of power begins to shift in favor of the other, weaker state? As the weaker state becomes stronger, it becomes increasingly dissatisfied with the existing distribution of benefits and begins to demand that the status quo be revised in its favor.

These demands confront the declining state with a clear, albeit unpleasant, choice if there is complete information. If the declining state knows the rising state's payoffs, the declining state also knows how much it has to concede in order to meet the rising state's minimal demands. The stark choice facing the declining state is either to make those concessions or fight. Asymmetric information makes the choice less stark but more complicated by creating a familiar trade-off. The more the declin-

ing state offers, the greater the chances of at least temporarily satisfying the rising state's demands, but the less satisfying the agreement will be.

As will be seen, the declining state deals with this trade-off by making a series of concessions. It concedes little to the rising state in the early phase when the rising state is still relatively weak and unlikely to be willing to use force. But as the rising state becomes stronger, the declining state grants more concessions. Throughout the shift in power, the declining state decides how much to concede by balancing the marginal gain of reducing the probability of war by offering slightly more with the marginal cost of having granted slightly more and therefore being a little less satisfied with any settlement. In general, the declining state's concessions are enough to satisfy the rising state if that state is relatively irresolute and unwilling to use force. In this case, the status quo is revised in favor of the rising state and the shift in power passes without war. However, the declining state will not concede enough to appease the rising state if that state is resolute and more willing to use force. In these circumstances, the shift in the distribution of power ends in war. And, the more willing the rising state is to use force, the sooner war comes.

These results clarify some existing claims about the effects of a shift in the distribution of power and contradict others. Drawing on Thucydides' ideas about the causes of the Peloponnesian War, Gilpin (1981) develops a theory of hegemonic war. (A hegemonic war is a war that determines which state will dominate the international system and thereby be able to define the international order and the distribution of benefits associated with that order.) He argues that different rates of economic growth produce shifts in the distribution of power among states. As a once weak state becomes stronger, it becomes increasingly dissatisfied with the existing distribution of benefits. Eventually the disparity between the distributions of benefits and power becomes so great that the rising state resorts to force to effect a more favorable distribution.

This argument explains one reason why the distribution of power may change, but it is less clear about why the declining state would wait for the rising state to become so dissatisfied that it would go to war. If the declining state would rather satisfy the rising state's demands than fight, why does it not try to do so by making concessions to the rising state in order to avert war? If, alternatively, the declining state would rather fight than satisfy the rising state's demands, why does it wait for the rising state to become stronger? Why does the declining state not attack the rising state while the latter is still relatively weak?

Viewing the shift in the distribution of power in terms of asymmetric information and the trade-off it creates provides an answer. Unsure of the rising state's willingness to use force and of how much it would take to satisfy its demands, the declining state does try to alleviate some of the mounting pressure by making a series of concessions. But the declining state is also willing to accept some risk of war and so does not offer enough to be certain of satisfying the rising state. Whether or not the shift in power ends in war depends on how willing the rising state actually is to use force.

We can see this pattern in the series of concessions Britain made to the United States during the nineteenth century as the distribution of power shifted in favor of the latter. In disputes over Maine, Texas, Oregon, Nicaragua, Venezuela, and other non-territorial issues, Britain came to accept the westward expansion of the United States and acquiesced in the rise of American power. When at century's end the United States wanted to build, fortify, and control a canal across the isthmus of Panama—all of which directly contravened the Clayton-Bulwer Treaty of 1850—Britain “granted every significant demand of the United States” (Allen 1955, 603). Although the British admiralty thought that construction of a canal, let alone the fortification and American control of it, was detrimental to British interests, Britain acceded to all of these things in the Hay-Pauncefote Treaty of 1901, which superseded the Clayton-Bulwer Treaty and “amounted to a major concession to the growth of American power” (Bourne 1967, 348; also see Friedberg 1988, 169–74).

Viewing the strategic problem confronting a declining state during a shift in the distribution of power from the perspective of asymmetric information also contradicts some of the claims of the power-transition school. That school argues that the most dangerous time during a shift in the distribution of power occurs at a transition, i.e., when the rising and declining states are roughly equal in power (Organski 1968; Organski and Kugler 1980). However, what the declining state concedes to the rising state at any point during a shift is determined by weighing the marginal gain of a lower probability of war against the marginal cost of a slightly less favorable agreement. As will be shown, whether the distribution of power is even or not has little bearing on these marginal calculations. Shifts in the distribution of power may be dangerous, but there is nothing especially dangerous about the point at which the rising and declining states are equally powerful.

A third way that a state can deal with threats in the shadow of power is by aligning with others, and chapter 5 examines this response. Do

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, S_1 , S_2 , and S_3 , and S_1 is threatening to attack S_2 . At a minimum, S_3 has three options. It can bandwagon by joining the threatening state S_1 , balance by aligning with S_2 against the would-be attacker, or wait by not joining either S_1 or S_2 .

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 S_1 and S_2 will result in an unfavorable shift in the distribution of power against S_3 . This latter factor encourages S_3 to enter the fray. Once S_3 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 S_3 more likely to join the stronger state. But there is an opposing factor. If S_3 aligns with the stronger state and that coalition prevails, then S_3 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 balancing and 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

¹³Dessler (1991) emphasizes the need to focus on causal mechanisms in international relations theory. Elster (1989) gives a brief overview of the role of mechanisms in the social sciences more generally, and Miller (1987) offers a more comprehensive treatment of mechanisms and causal theories.

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.¹⁵ 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.¹⁶ 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.¹⁷ 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.

¹⁵ Chapter 6 discusses some of the formal work that is already being done on domestic politics.

¹⁶ See Fearon (1995b) and chapter 3 below for a discussion of this.

¹⁷ The “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 (1995), 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 Altfeld'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 at the outset we “know” less than 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.²⁰

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

²⁰For 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 Gunnar Myrdal's rejection of efforts to formalize their ideas and what Krugman calls the “fall and rise of development economics.”

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.²¹

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

²¹Of course showing that a conclusion does not follow from a particular set of assumptions does not imply that the conclusion is wrong as a statement of empirical fact. States may be more likely to balance than bandwagon, or an even distribution of power may be more peaceful than a preponderance of power regardless of what a particular model says. If so, then, to the extent that the model also reflects the literature's current understanding of these issues, the model's failure also suggests that the understanding reflected in the literature is incomplete. This, indeed, is one of the ways that we learn from the modeling enterprise even when a model gets important things wrong. (As already noted, unfortunately it has been very difficult to establish many robust empirical regularities in international relations. As elaborated below, the empirical support for the claims that states generally balance or that a particular distribution of power is more peaceful is mixed at best.)

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.²² 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.²³ 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

²²Kreps (1990) provides an accessible overview of what game-theoretic work in economics has achieved, and much of what he says applies to political science as well.

²³See appendix 1 for a brief discussion of game trees and Morrow (1994b, 58–65) for a more extensive treatment.

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 that 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.²⁴

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

²⁴Although 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.

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.

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.