In the Shadow of Power

States and Strategies in International Politics

ROBERT POWELL

PRINCETON UNIVERSITY PRESS

PRINCETON, NEW JERSEY

making a series of concessions that "screen" the rising state according to its resolve. If the rising state is relatively unwilling to use force, these concessions satisfy its minimal demands, and the shift in power passes peacefully. If the rising state is more willing to use force, these concessions are not enough, and the shift ends in war. And, the more resolute the rising state, the earlier it attacks.

5

Alignment Decisions in the Shadow of Power

When one state threatens another, a third state has at least three options. It can align with the threatened state, align with the state making the threat, or try to avoid taking part in the conflict by waiting. All three behaviors are common and have a long history in international politics.

In May 1667, France began the War of Devolution by invading the Spanish Netherlands. Within a few months, England and the Dutch Republic had ended the war between them and joined with Sweden in the Triple Alliance to mediate a settlement between France and Spain and to use force against France to impose one if necessary. But, when the Dutch War began four years later, England and Sweden had switched sides by allying with France. Indeed, England was the first to declare war (McKay and Scott 1983, 20–36; Stoye 1969, 259–95).

In the spring of 1741, France, Bavaria, Prussia, and Spain bandwagoned together against a weakened Austria in the War of the Austrian Succession. Count Belle-Isle, who was, in effect, the French foreign minister, was trying to "create a coalition to break up Austria: the Southern Netherlands should go to France, Silesia to Prussia, Bohemia to Bavaria, and the Italian lands to Savoy and the Spanish princes." Britain, realizing that "Austria's existence was at stake and that her destruction would raise the power of France and her German allies to a level which would threaten Hanover and ultimately Britain herself" (McKay and Scott 1983, 165–66), then balanced by allying with Austria (Anderson 1995; Browning 1993; Thomson 1957).

A decade and a half later, the growing conflict between Britain and France in the New World triggered a "Diplomatic Revolution" on the Continent. A now stronger Austria allied with Russia to partition Prussia, and France subsequently joined in when Prussia preemptively attacked after the initial Austro-Russian attack had been delayed. Meanwhile Britain balanced with Prussia to keep French forces on the Continent and to protect Hanover (Horn 1957; McKay and Scott 1983, 181–96; Robson 1957).

States bandwagoned with stronger states, balanced against them, and tried to remain neutral throughout the French Revolutionary and Napoleonic Wars as well. Britain, for example, initially tried to remain neutral and then balanced against France from 1793 onward. Prussia fought against France in the First Coalition, let others do the fighting during the War of the Second Coalition, and subsequently fought with and against France until Napoleon's final defeat at Waterloo. Surveying this period, the diplomatic historian Paul Schroeder concludes that "every major power in Europe except Great Britain—Prussia, Austria, Russia, Spain—bandwagoned as France's active ally for a considerable period" (1994a, 121). Indeed, Austria's decision following Napoleon's disastrous invasion of Russia to join the Fourth Coalition with Britain, Prussia, and Russia can be seen as an effort to shape the final settlement by bandwagoning against France.[1]

Bandwagoning, balancing, and waiting were also common throughout the nineteenth and twentieth centuries. Britain and France allied against Russia in the Crimean War and were eventually supported by Austria while Prussia remained neutral. In 1866, France stood aside when Austria and Prussia went to war, and Austria reciprocated four years later by standing aside when France and Prussia went to war. In 1894, Republican France and Czarist Russia aligned against Germany, and Britain and France moved closer together after the turn of the century in the Entente Cordiale. After war broke out in 1914, the United States waited until 1917 before it entered the war. The United States also remained neutral in the Second World War until attacked by Japan. And Stalin, after trying to ally with Britain and France to oppose Hitler, joined with him in the partition of Poland.

Previous chapters examined how a state responds to threats through reallocating its internal resources or through bargaining and compromise. This chapter considers a state's efforts to deal with threats by aligning with other states and drawing on their resources. What factors affect a state's alignment decisions?

Is there a strong relation between the distribution of power and the choices states make, and, if so, can we explain this relation theoretically? Do states generally balance against the most powerful or most threatening state? Or do states bandwagon by aligning with the most

[1] For discussions of balancing, bandwagoning, and remaining neutral during this period, see Gulick (1955), Rosecrance and Lo (1996), Schroeder (1994a, 1994b), and Schweller (1994).

powerful or most threatening state? Or does a state's behavior vary depending on the circumstances facing it?

And, how does the technology of coercion affect states' alignment decisions? Suppose there are "increasing returns to scale in the aggregation of military capabilities" in the sense that the effective military capability of two aligned states fighting together is greater than the sum of the states' individual capabilities. Does this kind of increasing returns to scale make bandwagoning or balancing more likely?

This chapter probes these questions by extending the two-actor bargaining model developed in chapter 3 to a three-actor setting. The alignment game developed below investigates the relation between the distribution of power and states' alignment decisions in the simplest possible formal setting. Even more than in the previous chapters, the present analysis is exploratory and tentative, and is, at most, an early step in a modeling dialogue. As such, it advances our understanding in three ways.

First, specifying the model helps frame the problem and identify critical issues that any analysis—whether formal or not—would have to address. Second, even if states generally do balance as a matter of empirical fact (and this factual claim is contentious as elaborated below), several current arguments explaining why they do are inadequate. The model shows that the link these arguments make from assumptions to purported conclusions is at best incomplete. The analysis in this chapter thus illustrates Krugman's observation that "model-building, especially in its early stages, involves the evolution of ignorance as well as knowledge." (1995, 79). Third, the model suggests that the circumstances that produce balancing behavior are much more restrictive than has been previously appreciated.

Whether a state balances, bandwagons, or waits depends very much on the underlying technology of coercion and, especially, the extent to which there are increasing returns to scale in the aggregation of military capabilities. If these returns are small, i.e., if the effective military capability of two aligned states fighting together is about the same as the sum of the states' individual capabilities, then states generally prefer waiting to balancing or bandwagoning. That is, states tend to "pass the buck" (Christensen and Snyder 1990) to other states by letting them bear the cost of fighting when the returns to scale are slight.

Larger returns to scale undermine the strategy of waiting. The more that military forces cumulate, the more the distribution of power shifts against a state if it stands aside while others fight. This adverse shift makes waiting more costly relative to aligning and induces a state to join the conflict. This can be seen as a form of "chain-ganging" (Christensen and Snyder 1990) in which one state's attack on another compels a third state to enter the fray in order to prevent an adverse shift in the distribution of power.

A state faces a trade-off when deciding how to enter the conflict. A state makes itself relatively more powerful with respect to its coalition partner if it balances by aligning with the weaker of the two other states. This more powerful position reduces its vulnerability to its coalition partner and subsequently enhances that state's ability to secure its interests *if the coalition prevails*. This consideration would seem to make balancing more likely. But there is an opposing factor at work. A state maximizes its chances of being part of a winning coalition if it bandwagons by aligning with the stronger of the two other states. This consideration would seem to make bandwagoning more likely. The model developed in this chapter takes both of these opposing factors into account, and the latter consideration generally outweighs the former. States usually bandwagon if there are large returns to scale. Balancing does occur, however, if there are large returns to scale *and if* the attacker is much more willing to use force than the other states are.

The next section briefly surveys some of the existing work on the relation between alignment decisions and the distribution of power. The third section describes the model which, to keep the analysis manageably simple, focuses only on the case of complete information. As will be seen, a number of assumptions have to be made in order to specify the model and define the states' payoffs. The fourth section examines the payoffs to balancing, bandwagoning, and waiting to ensure that the combined effect of the individually plausible assumptions underlying these payoffs is also reasonable. There follows an analysis of a state's alignment decisions when the states are equally resolute. The sixth section then traces the feedback effects that this decision has on the would-be attacker's decision about whether and whom to strike. The final section reexamines the states' decisions when the attacker is more willing to use force than the other two states.

Balancing versus Bandwagoning

What is the relationship between a state's alignment decisions and the distribution of power among the states? Balance-of-power theory offers one answer to this question: When deciding on which coalition to join, a state generally chooses to align with the weaker of two coalitions. The combined effect of the states' individual decisions in turn produces a balance or roughly even distribution of power between the two coalitions.

The notion that rulers or states balance has a long history. David Hume ([1752]/1898) finds the idea in the writings of Thucydides and Xenophon and in the politics of ancient Greece. It was also present in Italy during the Renaissance: "In the 1440s there began to form in certain Italian minds a conception of Italy as a system of stable states, coexisting by virtue of an unstable equilibrium which was the function of statesmen to preserve" (Mattingly 1955, 71).[2] And, reflecting the then new Newtonian ideas of mechanics and equilibrium, "[T]he conception of a natural balance between states of unequal capabilities provided the foundation for early eighteenth-century theories of international relations. This idea is clearly present in the balance-of-power theories of the period" (Knutsen 1997, 121).[3]

Efforts to explain states' alignment decisions theoretically by explicitly deriving them from clear assumptions about states' ends and the strategic environment in which states pursue their ends are more recent. Indeed, constructing just such an explanation has been one of the major objectives of realist and, later, structural realist theories. As Waltz observes, "If there is any distinctively political theory of international politics, balance-of-power theory is it. And yet one cannot find a statement of the theory that is generally accepted" (1979, 117).

After making this observation, Waltz goes on to provide one of the clearest statements of the theoretical problem and attempts to solve it (116–28).[4] For him, balance-of-power theory "is a theory about the results produced by the uncoordinated actions of states" (122). The theory begins with "assumptions about the interests and motives of states" (122) and about the strategic setting in which those states pursue their interests. Based on those assumptions, the theory then makes predictions about the outcomes that emerge from the states' interaction in this environment.

[2] For a skeptical view of the existence of a coherent conception of balancing before the middle of the seventeenth century, see Butterfield (1966, 139).

[3] See Haas (1953), Hinsley (1963), and Knutsen (1997) for historical overviews of balance-of-power theories and thinking.

[4] For earlier realist efforts to construct a balance-of-power theory, see Gulick (1955), Kaplan (1957), and Morgenthau (1967). Haas (1953), still earlier, takes an important initial step by helping to clear the theoretical ground by systematically describing the different ways that the term "balance of power" had been used.

Waltz's formulation of balance-of-power theory is very spare. Indeed, he argues that two assumptions are sufficient to explain the tendencies of balances of power to form. The first pertains to the motives of the units. The states are assumed to be "unitary actors who, at a minimum, seek their own preservation and, at a maximum, drive for universal domination" and which "try in more or less sensible ways to use the means available in order to achieve the ends in view" (1979, 118). The second assumption characterizes the environment in which the states seek their ends. It is that the system is anarchic, i.e., there is "no superior agent to come to the aid of states that may be weakening or to deny to any of them the use of whatever instruments they think will serve their purposes" (118). According to Waltz, these two assumptions lead to the "recurrent formation of balances of power" (119). "Balance-of-power politics prevail whenever two, and only two requirements are met: that the order be anarchic and that it be populated by units wishing to survive" (121).

Believing that balances of power "often fail to form," Stephen Walt offers what he calls "balance-of-threat theory" as a refinement of balance-of-power theory (1988, 279–82; 1987). States respond to imbalances of power in balance-of-power theory, and balancing and bandwagoning are defined in terms of power. To wit, a state balances if it joins the weaker of two coalitions and bandwagons if it aligns with the stronger coalition. Walt, however, argues that states respond to threats rather than capabilities alone, and he redefines balancing and bandwagoning accordingly. A state balances if it joins the less threatening coalition and bandwagons if it aligns with the more threatening coalition in an effort "to appease it or profit from its victory" (1988, 278). Based on a study of alliances in the Middle East from 1955 to 1979, Walt (1987) then concludes that states generally balance against threats.[5]

The claim that states tend to balance has been criticized recently on two scores. The central thrust of Randall Schweller's criticism is that balance-of-threat theory focuses only on

> cases in which the goal of alignment is security, and so systematically excludes alliances driven by profit. Yet, as Walt himself claims, one of the primary motivations for bandwagoning is to share in the spoils of victory. When profit rather than security drives alliance choices,

[5] Walt (1988) subsequently examines the cases of Iran, Turkey, India, and Pakistan during the cold war and claims that these cases also support the hypothesis that states balance against threats.

> there is no reason to expect that states will be threatened or cajoled to climb aboard the bandwagon; they do so willingly.... Thus we will not observe cases of bandwagoning for profit by examining alliances as a response to threats. We must look instead at alliance choices made in the expectation of gain. (1994, 79)

Schweller in effect believes that Walt's cases suffer from a selection bias (King, Keohane, and Verba 1994). That is, Walt's sample of cases is biased toward finding balancing to be more common than bandwagoning. This bias arises because Walt focuses on a set of cases in which the states deciding how to align are under direct threat rather than focusing on a broader and more representative set of cases in which the aligning states might bandwagon in order to share in the spoils of victory. Based on this analysis and a brief review the Italian Wars of 1494–1517, the French bid for hegemony in 1667–79, and the Napoleonic Wars, Schweller concludes that bandwagoning is more prevalent than Walt believes (1994, 93).[6]

Paul Schroeder's criticism centers on balance-of-power theory. He asks whether the theory "is adequate and useful as an explanatory framework for the history of international politics in general, over the whole Westphalian era from 1648 to 1945" (1994a, 110). Based on his reading of history, he concludes that balance-of-power theory fails in this task (1994a, 1994b). The theory "is incorrect in its claims for the repetitiveness of strategy and the prevalence of balancing in international politics" (1994a, 120). Indeed, numerous historical examples "make a *prima facia* case" against these claims (1994a, 120). Overall, Schroeder sees "bandwagoning as historically more common than balancing" (1994a, 117).

This brief review of the conflicting theoretical and empirical claims about the relative prevalence of balancing and bandwagoning suggests that the answer to the question of whether states generally balance or bandwagon is that it depends on the circumstances in which they find themselves. For if this decision did not depend on circumstances and if there were a strong tendency to act in one way or the other, then this pattern should have been readily apparent in the historical record. We would have several large-scale statistical and comparative case studies demonstrating this tendency.

But suggesting that a state's alignment depends on circumstances is to beg the most important question. On what circumstances does it depend and does this decision vary systematically with changes in these

[6] See Labs (1992), Kaufman (1992), and Walt (1992) for additional critiques of Walt's analysis and his response.

circumstances? The model developed in the next section takes a step toward answering this question by examining the effects of the distribution of power and the technology of coercion on a state's alignment decisions. Ideally, the analysis will give us a clearer and sharper idea of what patterns we should expect to see in the historical data and experience.

A Model of Alignment

The specification of a model is guided by the questions it is intended to help answer. Three questions are of primary concern here. What is the relation between the distribution of power and a state's alignment decision? How does this decision affect whether or not a dissatisfied state actually attacks and, if so, which state it attacks? And, how do changes in the technology of coercion affect these choices?[7]

In order to address these issues, the model needs to include at least three features. First, one state must have the option of deciding whether or not to attack, and, if it does decide to strike, it must also be able to select which state or states will be attacked. Second, if there is an attack, then the state that is *not* attacked must have the choice of bandwagoning with the attacker, balancing by aligning against the attacker, or waiting.[8] Third, the states' payoffs to these different choices should depend on the distribution of power among the states and on the technology of coercion. A fourth feature would also be desirable. It would be nice if the model grew naturally out of the models considered in previous chapters. Such a connection would bring both technical and substantive benefits.

[7]Other game-theoretic efforts to study the relation between the distribution of power and states' alignment decisions include Wagner (1986), who renewed interest in the problem, and Niou and Ordeshook (1990, 1991). Kaplan, Burns, and Quandt (1960) provide an early game-theoretic discussion of this problem.

A major difference between the model developed below and Niou and Ordeshook's is the technology of coercion. They assume what might be called a "voting" technology in which a state is *certain* to prevail over another even if the former is only slightly larger than than the latter, i.e., if the former has slightly more than 50 percent of the resources. Thus a larger state is always dissatisfied with a smaller state. By constrast, the present model assumes that a state that has only slightly more military resources or capabilities than another state is also only slightly more likely to prevail in the event of war.

[8]Unless otherwise noted, the term "bandwagoning" is used throughout the rest of this chapter to mean aligning with the attacking state and "balancing" is used to mean aligning with the attacked state. This usage is more akin to the way the these terms are used in the balance-of-threat analysis than in the balance-of-power argument. This usage is, however, purely a matter of convenience, and we will discuss below whether or not states generally bandwagon or balance and, if so, whether it is against power or threats.

Technically, it might be possible to use the solutions to the previous models to help solve the model of alignment decisions. Substantively, the more closely related the alignment model is to previous models, the easier it is to see the effects of moving from the two-actor setting in those models to the multiple-actor setting of the alignment model.

These features can be captured most simply in a three-actor, two-stage game. Let these three states be called A, S_1, and S_2, and suppose that, as in past models, there is an initial distribution of power among these states as well as an initial distribution of benefits. The distribution of power is discussed at length below. For now, let the probability that state j defeats state k if these two states fight each other and the third state stays out of the conflict be denoted by p_k^j. (The key to this notation, which is employed throughout this chapter, is that $p_{\text{subscript}}^{\text{superscript}}$ denotes the probability that the state or coalition in the superscript defeats the state or coalition in the subscript.) As in previous models, war always eliminates one of the states, so the probability that k defeats j is just $p_j^k = 1 - p_k^j$. Then the initial distribution of power can be described in terms of three probabilities: the probability that the potential attacker, A, would defeat S_1 if they fought in isolation; the probability that the potential attacker would defeat S_2 if they fought in isolation; and the probability that S_1 would defeat S_2 if they fought in isolation. These three probabilities are denoted by p_1^A, p_2^A, and p_2^1.

The distribution of benefits may be thought of as the amount of territory each state initially controls. If, for example, the states initially control the same amount of territory, then the distribution is $\frac{1}{3}$ for the first state, $\frac{1}{3}$ for the second, and $\frac{1}{3}$ for the third where, by assumption, the total amount of territory is one. More generally, the initial distribution of territory can be represented by q_A, q_1, and q_2 which are, respectively, the amount of territory A, S_1, and S_2 control and where $q_A + q_1 + q_2 = 1$.

The first stage of the game models the states' alignment decisions and whether or not one of the states attacks given the initial distributions of power and benefits. If there is no attack in this stage, the game ends peacefully. If there is an attack, then at least one of the states is eliminated as was the case in previous models. This elimination transforms the three-actor game into a two-actor game and begins the second stage. This two-actor, second stage is modeled like the two-actor bargaining game studied in chapter 3.

Figure 5.1 illustrates the sequence of moves in the first stage. The game begins at the open dot where the potential attacker, A, must

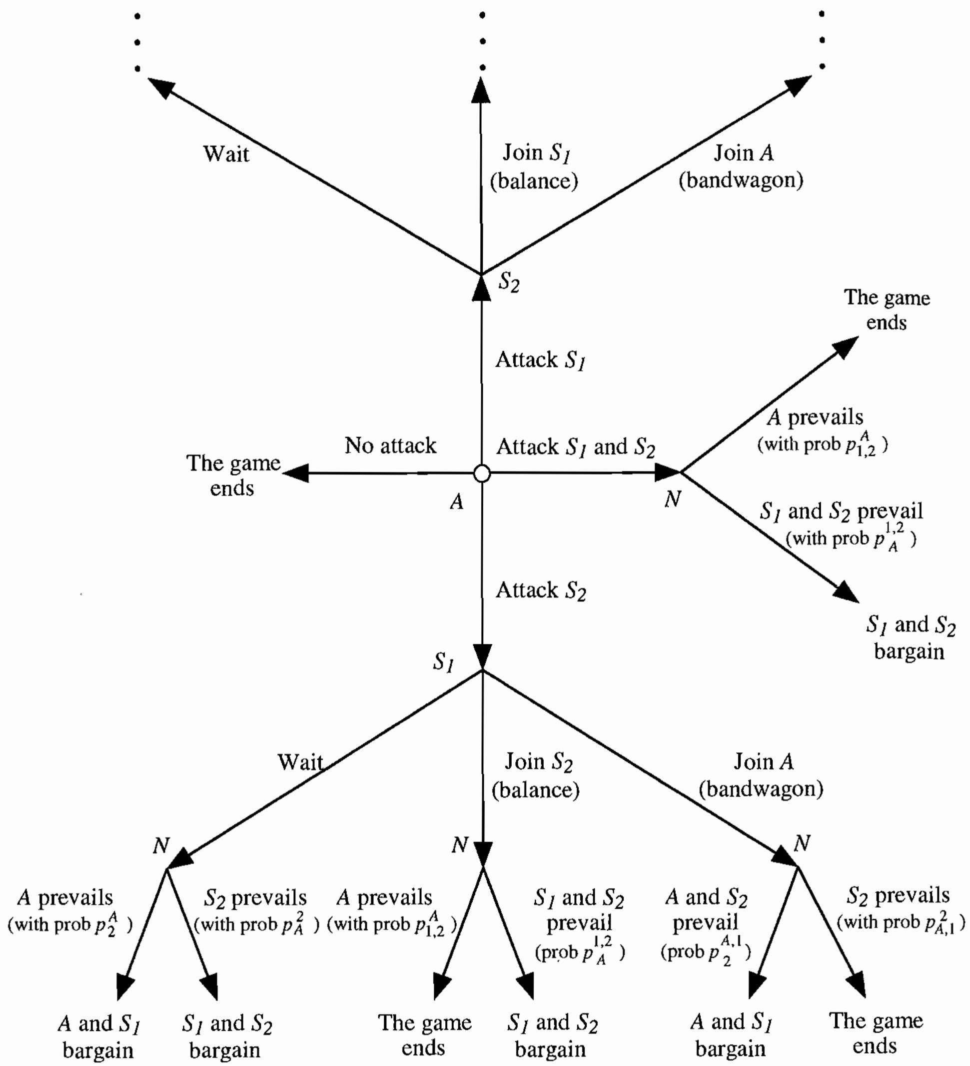


Figure 5.1 The alignment game

choose from four alternatives. It can decide not to attack anyone, in which case the game ends. Or, A can attack S_1 alone, S_2 alone, or attack both S_1 and S_2 at once. If A attacks both states, then either A prevails and eliminates its adversaries, in which case the game ends because there is only one state left, or the coalition of S_1 and S_2 eliminates A and the second bargaining stage begins.

If A attacks only one of the other two states, then the state that has not been attacked must decide what to do. For example, suppose that A attacks S_2. Then S_1 can wait by letting A and S_2 fight it out by themselves; S_1 can balance by aligning with S_2 to oppose the attack; or S_1 can bandwagon by joining with A in the attack.

If S_1 waits, then A prevails and eliminates S_2 with probability p_2^A, and S_2 prevails with probability $p_A^2 = 1 - p_2^A$. In the game tree, this uncertain outcome is represented by a random move made by a player called N (for "Nature"), which plays the branch "A prevails" with probability p_2^A and the branch "S_2 prevails" with probability p_A^2. If A does prevail, then the first stage ends in the elimination of S_2 and the second stage begins with A and S_1 bargaining with each other. If, by contrast, S_2 prevails, then A is eliminated and the second stage begins with S_1 and S_2 facing each other.

If S_1 bandwagons with A, then either the coalition of A and S_1 prevails by eliminating S_2, or S_2 prevails by eliminating A and S_1. If the coalition prevails—which it does with probability $p_2^{A,1}$ where, recall, $p_{\text{subscript}}^{\text{superscript}}$ denotes the probability that the state or coalition in the superscript defeats the state or coalition in the subscript—then the second round begins with A and S_1 confronting each other. If, by contrast, S_2 prevails, the game ends with S_2 in control of all of the territory.

Finally, if S_1 balances by joining S_2, then the coalition of S_1 and S_2 prevails with probability $p_A^{1,2}$. In that event, S_1 and S_2 go on to the second stage. If, however, A prevails, then the game ends and there is no second round as A controls all of the territory and faces no other states.

The second stage begins once one of the states has been eliminated, and what happens in this stage is completely analogous to the complete-information version of the bargaining game studied in chapter 3. The elimination of one of the states in the first stage results in a new distribution of power and a new distribution of benefits between the two remaining states, and these two states bargain about revising this distribution of benefits during the second stage. In particular, one of the states can propose a new territorial division, which the other state can accept, counter with another proposal, or reject in order to impose a new distribution. If the second state accepts, the territory is divided as agreed and the game ends. If the second state attacks, then one or the other of the states is eliminated and the game ends. If the second state makes a counter-offer, then the first state can accept, attack, or make a counter-proposal and so on as in the model described in chapter 3. (The second stage of the alignment game is defined formally in appendix 5.)

Four points should be made about this tree. First, states decide what to do in the first stage based on their expectations about how these decisions will ultimately affect the outcome of the second stage. If, for example, a state believes that it will have to make significant concessions during the bargaining round if it waits while others fight in the first round, then this state is less likely to wait and more likely to align with one of the other states. The bargaining round, therefore, has an important effect on the states' alignment decisions even though it comes after these decisions have been made.

The second observation is that the bargaining stage is closely related to the principle of compensation according to which a victim's territory is divided so "as not to change decisively the strength of any victor in relation to his partners" (Mattingly 1955, 141). Mattingly argues that this principle played an important role in Renaissance diplomacy: "In the arrangements for cutting up the Milanese between France and Venice, or Naples between France and Spain, or the Venetian territories among the allies of the League of Cambrai, the principle was more or less consciously observed (141)." Similarly, Morgenthau (1967, 173) observes that the Treaty of Utrecht (1713), which ended the War of Spanish Succession, divided Spain's possessions between Austria and France so as to conserve the balance of power. The principle was also at work at the Congress of Vienna at the end of the Napoleonic Wars (Gulick 1955) and most dramatically in the partitions of Poland in 1772, 1793, and 1795:

> Since territorial acquisitions at the expense of Poland by any one of the interested nations—Austria, Prussia, and Russia—to the exclusion of the others would have upset the balance of power, the three nations agreed to divide Polish territory in such a way that the distribution of power among themselves would be approximately the same after the partitions as it had been before. (Morgenthau 1967, 173)

Something like this happens in the bargaining stage of the model. As shown in chapter 3, the bargaining results in a distribution of benefits that mirrors the underlying distribution of power. Thus, the victim's territory is divided between the victors in proportion to the distribution of power between the victors. But this division does not reflect a principle in the model, it reflects the underlying distribution of power.

The third point to be made about the game tree is that the model focuses on intra-war alignment decisions. S_1, for example, decides to wait, balance, or bandwagon only after A has decided to attack. This focus is, however, broader than it might first seem. If, for example, S_1 and

S_2 had previously formed an alliance (something not explicitly modeled), then S_1's first-stage decision could be seen as a choice between honoring an alliance or reneging on it.

The analysis of intra-war alignment decisions is, moreover, a prerequisite to analyzing prewar alignment decisions. Just as the states' expectations about the bargaining stage affect the alignment decisions they make in the first stage of the game, the states' expectations about what others will actually do in the event of war affects their peacetime alignment decisions. As noted at the beginning of the chapter, the game is an early step in the modeling enterprise.

The fourth observation about the tree is that the current analysis focuses on alignment decisions and not specifically on alliances, which are usually defined as "a subset of alignments—those that arise from or are formalized by an explicit agreement, usually in the form of a treaty" (G. Synder 1997, 8). The model does not explicitly address the important question of why states sign "scraps of paper" (G. Snyder 1997, 9). How and to what extent formal alliances can create credible commitments is discussed briefly and informally below.[9]

Now consider the state's payoffs. Three kinds of assumptions have to be made in order to specify the payoffs. These assumptions will be discussed informally first and then more formally at the end of this section.

The first assumption pertains to the basic goals of the states. As in the models discussed in chapters 3 and 4, each state derives benefits from the amount of territory it controls. Each state therefore tries to maximize the total amount of territory under its control.

The second kind of assumption describes the cost of fighting. A state's cost of fighting is assumed to reflect two factors: that state's general willingness to use force and the size of its adversary. In particular, the more resolute a state or the smaller its adversary, the lower that state's cost to fighting that adversary. Suppose, for example, that the potential attacker A fights S_1 in isolation, then A's cost is taken to be $c_A \cdot q_1$ where c_A represents A's marginal cost to fighting and q_1 is the size of its opponent S_1. This expression formalizes the idea that the more resolute the potential attacker (i.e., the smaller c_A) or the smaller an adversary's size (i.e., the smaller q_1), then the lower A's cost to fighting.[10]

[9] Other informal discussions of the issue are Stein (1990, 151–69) and G. Snyder (1997, 9–11). Fearon (1994), Morrow (1994a), and Smith (1998) model these issues formally.

[10] Allowing the cost of fighting to depend on the size of the adversary provides a richer formulation than that considered in the bargaining models in chapters 3 and 4. The effects of including this richer but also more complicated formulation in those models is discussed in chapter 6.

To specify the costs of fighting in and against a coalition, suppose that the potential attacker fights the coalition of S_1 and S_2. Then, A's cost is taken to be $c_A(q_1 + q_2)$ where the sum $q_1 + q_2$ is the combined size of A's adversary. If, instead, A and S_2 fight in a coalition against S_1, then A's cost is assumed to depend on its willingness to fight, the size of the third state S_1, and on the distribution of power between the members of the coalition. In particular, A's cost of fighting with S_2 against S_1 is $p_2^A \cdot c_A \cdot q_1$. This expression is just the cost A would bear if it alone fought S_1, namely $c_A \cdot q_1$, weighted by the distribution of power between A and S_2. This cost increases as A becomes more powerful relative to its coalition partner (p_2^A increases) and presumably bears a larger share of the fighting; decreases as A becomes more willing to use force (a lower c_A); and increases with the size of the third state q_1.

Finally, the third type of assumption needed to specify the states' payoffs describes the ways that various alignments affect the distribution of power. Indeed, three questions about the distribution of power must be answered in order to complete the specification:

(i) If two states align, what is the distribution of power between the coalition and a the third state?
(ii) If two states fight together in a coalition and are victorious, what happens to the distributions of power and benefits between them as a result of the fighting?
(iii) If one state waits while two others fight, what will be the resulting distributions of power and benefits between the victor and the state that waited?

The answer to the first question affects a state's payoff to bandwagoning and balancing by specifying the probabilities that it will be on the winning side if it aligns with the stronger or weaker state. The answer to the second question also helps determine a state's payoff to bandwagoning and balancing by describing the distributions of power and benefits that will exist between the coalition partners at the end of the first round of the game if their coalition prevails. These distributions thus determine which state, if either, will be dissatisfied at the outset of the second-stage bargaining and which state will have to make concessions to its former coalition partner. Finally, the answer to the third question essentially defines the payoff to waiting by characterizing the distributions of power and benefits that obtain if a state waits.

Although the accounting mechanism of the model forces these three questions to the surface, it is hard to believe than any systematic analysis

of the relation between states' alignment decisions and the distribution of power—whether formal or not—could successfully avoid these issues. As just noted, the answers to these questions describe how the states' decisions affect the distribution of power and, ultimately, the states' payoffs. The model does not make these questions important; it only makes them explicit and compels us to address them. Unfortunately, we do not have well-established empirical or theoretical answers to these questions, and this is an important task for furture work. We will proceed here by making analytically simple and, we hope, substantively fruitful assumptions.[11]

The first question really asks to what extent do military forces "add up"? How, that is, does forming a coalition affect the probability that the coalition will prevail? The basic idea behind the formulation used here begins by assuming that each state has a latent military capability at the outset of the game. Then the likelihood that one state prevails against another if they fight in isolation is the ratio of that state's military capability to that of its adversary. If, for example, one state's capability is twice that of another, then the odds that the former would prevail are two-to-one. Thus, the states' underlying military capabilities determine the initial distribution of power p_1^A, p_2^A, and p_2^1.[12]

Generalizing this idea, the likelihood that a coalition of two states defeats a third state is the ratio of the coalition's capability to that of the third state. The coalition's capability in turn depends on the individual capabilities of the member states and on the technology of coercion, which describes the degree to which military capabilities "add up." If, for example, there are *increasing returns to scale in the aggregation of military capabilities*, then a coalition's capability is larger than the sum of the individual capabilities of the members of the coalition. If there are

[11] As observed in chapter 1, one of the benefits of a modeling dialogue is that it helps make critical issues and assumptions explicit so that they can be more readily discussed and subjected to empirical and theoretical scrutiny.

[12] This way of specifying a state's power is typically used in quantitative work in international relations based on the Correlates of War data (see Singer and Small 1972; Small and Singer 1982) where a state's relative capability is generally taken to be a weighed average of its shares of total and urban population, military personnel and spending, and energy consumption along with iron and steel production. Accordingly, the use of relative capabilities as a measure for power in the present model not only provides a simple and consistent way of describing the effects of alignment decisions on the probability of prevailing, but it also makes it easier to relate the results derived from the model to empirical work. For a cautionary study that compares perceptions of Russian power before the First World War with estimates based on Correlates of War data, see Wohlforth (1987).

constant returns to scale, a coalition's capability is just equal to the sum of the members' capabilities. And if there are *decreasing returns to scale*, a coalition's capability is less than the sum of the members' capabilities. As will be seen, whether there are increasing, decreasing, or constant returns to scale has an important effect on states' alignment decisions. The parameter g is used to represent the degree to which military forces cumulate. As will be shown below, if there are decreasing, constant, or increasing returns to scale, then g is less than, equal to, or greater than one.

The answer to question (ii) above defines what the distributions of power and benefits are between the victors immediately after the defeat of the third state. These distributions specify the conditions under which the bargaining in the second stage takes place and makes it possible for a state to determine what its payoff would be if it survives into the second round. As we saw in chapter 3, if the distribution of power between the victors mirrors the distribution of benefits between them, then both states are satisfied and there are no further changes in the territorial status quo. If, by contrast, the distribution of benefits following the elimination of the third state does not reflect the distribution of power, then one of the victors is dissatisfied. In these circumstances, the satisfied state meets the dissatisfied state's minimal demands during the complete-information bargaining by conceding just enough territory to the dissatisfied state to make the dissatisfied state indifferent between attacking and accepting the concession.

A very simple assumption defines the initial distributions of power and benefits between the members of a winning coalition. The fighting involved in defeating and eliminating the third state is assumed to leave each of the victors in possession of an amount of the defeated state's territory proportional to the distribution of power between the victors. If, for example, one member of a coalition is twice as strong as the other, then the fighting ends with the former in control of twice as much of the defeated state's territory. It is important to emphasize that this is the distribution of territory resulting *directly* from the fighting, and it is subject to revision during the second stage of bargaining. Indeed, that is what the victors bargain about.

The distribution of power between the members of a coalition following the defeat of a third state is taken to be the same as it was before the war. That is, forming a coalition and fighting together effectively "freezes" the distribution of power between the members of the coalition. If, therefore, one member of a coalition is twice as strong as the other member when the coalition forms, then the former will also be twice as strong as the latter if the coalition fights and prevails.

Finally, suppose that one state attacks another and the third state stands aside while the other two fight. Because it waited, the third state is sure to survive into the bargaining stage. The answer to question (iii) defines the distributions of power and benefits that exist between the third state and the victor at the outset of the bargaining between them.

Suppose more concretely that A attacks S_2 and S_1 waits. The distribution of benefits between S_1 and the victor of the struggle between A and S_2 seems straightforward. By waiting and letting A and S_2 fight it out alone, S_1 neither gains nor loses any territory. Thus, S_1 controls q_1, and the victor controls $q_A + q_2$. What to assume about the distribution of power is, however, less clear. Once the victor has consolidated both A's and S_2's military capabilities, should we assume it to be stronger, weaker, or equal in strength to a coalition composed of A and S_2? To keep things as simple as possible, we assume that the victor in a fight between A and S_2 is equal in strength to the coalition formed between A and S_2. (The consequences of relaxing this assumption are discussed below.)

The remainder of this section describes the preceding assumptions about probabilities and payoffs in more detail and formalizes some of the payoffs. (Appendix 5 characterizes all of the payoffs.) Readers less interested in these details and formalities may want to skip the remainder of this section.

As a first step, it will be helpful to conceive of each state in terms of the territory it controls and its military capability. The territorial distribution is represented by q_A, q_1, and q_2, and the states' capabilities are denoted by k_A, k_1, and k_2. These capabilities define the probabilities that a state or coalition prevails. Suppose that two states, say A and S_1, fight in isolation. Then the probability that A prevails is the ratio of its capability to the total of A's and S_1's capabilities. That is, the probability that A prevails, p_1^A, is $k_A/(k_A + k_1)$.

Now suppose that the potential attacker and S_1 form a coalition and fight S_2. The probability that the coalition prevails, $p_2^{A,1}$, is the ratio between the coalition's military capability, which is $g(k_A + k_1)$, and the total capability of the coalition's and S_2's capabilities. This gives

$$p_2^{A,1} = \frac{g(k_A + k_1)}{g(k_A + k_1) + k_2}$$

where the parameter g measures the extent to which military forces "add up."[13] There are constant returns to scale if $g = 1$, because the coalition's capability is equal to the sum of the states' individual capabilities, i.e., $g(k_A + k_1) = k_A + k_1$. There are increasing returns to scale if $g > 1$, because the coalition's capability is greater than the sum of the states' individual capabilities. And there are decreasing returns if $g < 1$, because the coalition's capability is less than the sum of the members' separate capabilities.

To illustrate the way that the states' payoffs are specified, it is useful to consider some specific paths through the game tree in figure 5.1. Suppose A attacks S_2 and that S_1 then bandwagons by aligning with A and joining in the attack. The specification of S_2's payoff is straightforward because it fights alone and the game ends if it prevails. Accordingly, S_2's payoff to fighting the coalition is its payoff if it prevails times the probability of prevailing plus its payoff if it is defeated weighted by the probability of losing. Prevailing brings S_2 control over all of the territory and a payoff of one less the cost of fighting $c_2(q_A + q_1)$ where, recall, c_2 reflects S_2's willingness to fight and $q_A + q_1$ is the size of its adversary. In symbols, S_2's payoff to prevailing is just $(1 - c_2(q_A + q_1))$. Defeat, however, means that S_2 loses control over all of its territory and pays the cost of fighting. Thus, S_2's expected payoff to fighting the coalition of A and S_1 is $p^2_{A,1}(1 - c_2(q_A + q_1)) + (1 - p^2_{A,1})(0 - c_2(q_A + q_1))$, which reduces to $p^2_{A,1} - c_2(q_A + q_1)$ where $p^2_{A,1}$ is the probability that S_2 prevails over the coalition of A and S_1.

A's and S_1's payoffs depend on the probability that their coalition prevails and on whether the distribution of territory following the elimination of S_2 subsequently changes during the bargaining between A and S_1. If the coalition prevails, then the distribution of power following the conflict is assumed to be the same as it was before the defeat of S_2. Accordingly, the distribution of power at the start of the bargaining stage is the probability that A would defeat S_1 or p^A_1. As for the distribution of benefits between A and S_1 at the start of bargaining phase, each of the victors obtains a share of the defeated state's territory proportional to the distribution of power between them. The attacker A receives $p^A_1 q_2$ of S_2's territory, and S_1 obtains $(1 - p^A_1)q_2$.

What happens during the bargaining between A and S_1 depends on the distributions of power and benefits between them. Recall that a state

[13] This is just one of many possible formulations. For a discussion of some of the alternatives, see Hirshleifer (1988).

is dissatisfied if its expected payoff to fighting is greater than its payoff to living with the territory it controls. Both states, moreover, cannot be dissatisfied. As shown in chapter 3, at least one of the states has to be satisfied.

If both states are satisfied, i.e., if the distribution of power following the defeat of S_2 roughly mirrors the distribution of benefits, then neither A nor S_1 can credibly threaten to use force to revise the territorial status quo. The bargaining phase therefore ends with A controlling an amount of territory equal to its original share of the territory, q_A, plus what A captures from S_2, which is $p^A_1 q_2$. Similarly, the game ends with S_1 in control of $q_1 + (1 - p^A_1)q_2$.

If, by contrast, one of the states, say A, is dissatisfied, then S_1 concedes just enough territory during the bargaining stage to make A indifferent between accepting the concession or attacking. As in chapter 3, A's payoff to attacking equals its probability of prevailing, p^A_1, less its cost of fighting S_1, $c_A(q_1 + (1 - p^A_1)q_2)$. Accordingly, the bargaining ends with A in control of an amount of territory equal to $p^A_1 - c_A(q_1 + (1 - p^A_1)q_2)$ and S_1 in control of the remainder.

To illustrate the combined effects of the first and second stages, suppose that A attacks S_2, S_1 subsequently aligns with A, and A and S_1 would be satisfied with each other if their coalition prevails. Then A's payoff to attacking is the payoff A would receive during the bargaining phase weighted by the probability that the coalition prevails plus A's payoff to losing (which is zero minus its share of the cost of fighting S_2) weighted by the probability that the coalition is defeated. In symbols, A's payoff is $p^{A,1}_2(q_A + p^A_1 q_2 - c_A p^A_1 q_2) + (1 - p^{A,1}_2)(0 - c_A p^A_1 q_2)$, which simplifies to $p^{A,1}_2(q_A + p^A_1 q_2) - c_A p^A_1 q_2$. The first term of this simpler expression is just the amount of territory A controls if the coalition prevails weighted by the probability that the coalition prevails. The second term is A's share of the cost of fighting S_2.

Now suppose that A would be dissatisfied with S_1 after their coalition prevails. Then A's payoff to attacking S_2 if S_1 subsequently aligns with A is its bargaining-stage payoff weighted by the probability of surviving the first stage less the cost of fighting. In symbols, this is $p^{A,1}_2[p^A_1 - c_A(q_1 + (1 - p^A_1)q_2)] - c_A p^A_1 q_2$.[14]

[14] In order to simplify the analysis, these payoffs focus on the territory A obtains during the bargaining phase and disregard the territory A controls during the first phase. In effect, then, these payoffs ignore the role of discounting.

The Payoffs to Balancing, Bandwagoning, and Waiting

As the previous section made clear, many crucial assumptions have to be made in order to specify the states' payoffs to balancing, bandwagoning, and waiting. There are, moreover, several plausible alternatives for each of these assumptions, and, unfortunately, the existing theoretical and empirical work in international relations provides little or no guidance in choosing among these possibilities. The choices described here were made with an eye toward both substantive richness and analytic simplicity. Seen in this light, each assumption seems plausible when considered individually. But the payoffs to balancing, bandwagoning, and waiting combine these individual assumptions in different ways, and this poses a question. Do these combinations also seem plausible? This section examines the gross behavior of these payoffs to ensure that it, too, seems reasonable. Once this has been done, the next section will compare these payoffs to each other in order to determine how the states align.

We will use a simple comparative-static analysis to study the behavior of the individual payoffs and, later, the states' alignment decisions. This analysis begins at the point at which the distributions of power and benefits mirror each other. Then it traces the effects of an increase in the potential attacker's military capability while the other two states' military capabilities remain constant. This change increases the potential attacker's strength relative to each of the other states while leaving the relative strength between the other two states unchanged.

To specify the comparative-static analysis more precisely, we first need to identify the starting point of the analysis, i.e., the distribution of power that mirrors a particular territorial distribution. Recall that the distribution of territory is q_A, q_1, and q_2 where the potential attacker controls q_A, S_1 controls q_1, and S_2 controls q_2 with $q_A + q_1 + q_2 = 1$. Then, just as in the model examined in chapter 3, the distribution of power between two states, say A and S_1, mirrors the distribution of benefits between them if the probability that A defeats S_1 equals A's share of the total territory that these states control. That is, the distributions of power and benefits between A and S_1 mirror each other if $p_1^A = q_A/(q_A + q_1)$. Similarly, the distribution of power between A and S_2 and between S_1 and S_2 is the same as the distribution of benefits between them if $p_2^A = q_A/(q_A + q_2)$ and $p_2^1 = q_1/(q_1 + q_2)$. Thus we can find the distribution of power that mirrors any initial distribution of territory, and this is where the comparative-static analysis starts.

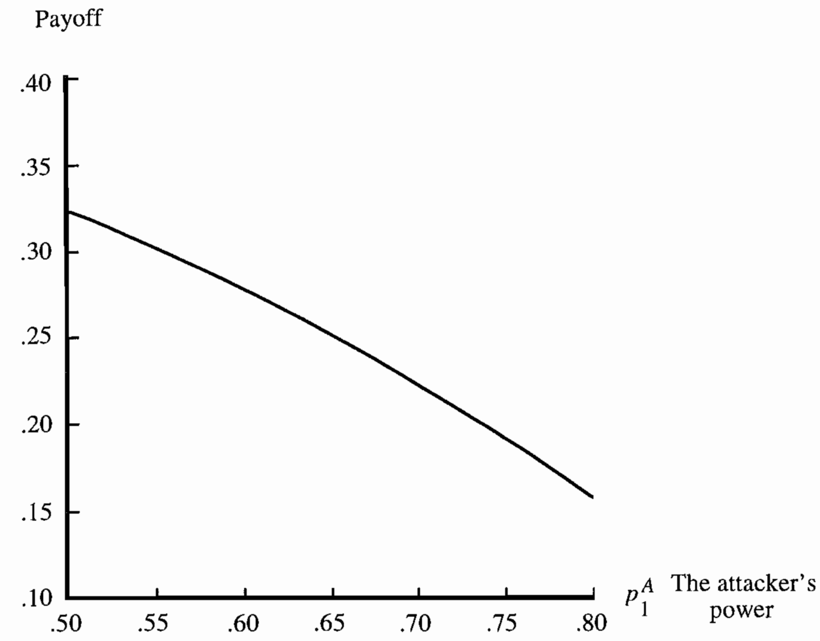


Figure 5.2 The payoff to balancing

As the potential attacker's military capability grows, it becomes relatively more powerful, and the probability that it would prevail over S_1 in a bilateral contest rises. In symbols, p_1^A increases as A's capabilities grow. Similarly, the probability that A would prevail over S_2, p_2^A, also increases. However, the distribution of power between S_1 and S_2, p_2^1, remains constant because there has been no change in these states' capabilities.

Figure 5.2 illustrates the effects of an increase in A's capabilities on the payoff to balancing given that each state initially controls an equal share of the territory ($q_A = q_1 = q_2 = \frac{1}{3}$). More specifically, the figure plots S_1's payoff to aligning with S_2 given that A has attacked S_2.[15] Intuitively, we would expect the payoff to balancing against a stronger attacker to be less than the payoff to balancing against a weaker attacker, and this is precisely what the figure shows. Starting at $p_1^A = \frac{1}{2}$ where the distribution of power mirrors the distribution of benefits, i.e., where

[15] S_1's and S_2's payoffs to balancing are the same because these states are identical in size, power, and willingness to use force. Thus, figure 5.2 also depicts S_2's payoff to balancing with S_1 against A if A attacks S_1.

$p_1^A = q_A/(q_A + q_1) = (\frac{1}{3})/(\frac{1}{3} + \frac{1}{3}) = \frac{1}{2}$, the payoff to balancing steadily declines as A's military capability grows and p_1^A increases. (It will be convenient to represent the changes in the potential attacker's military capabilities implicitly by plotting p_1^A along the horizontal axis.)

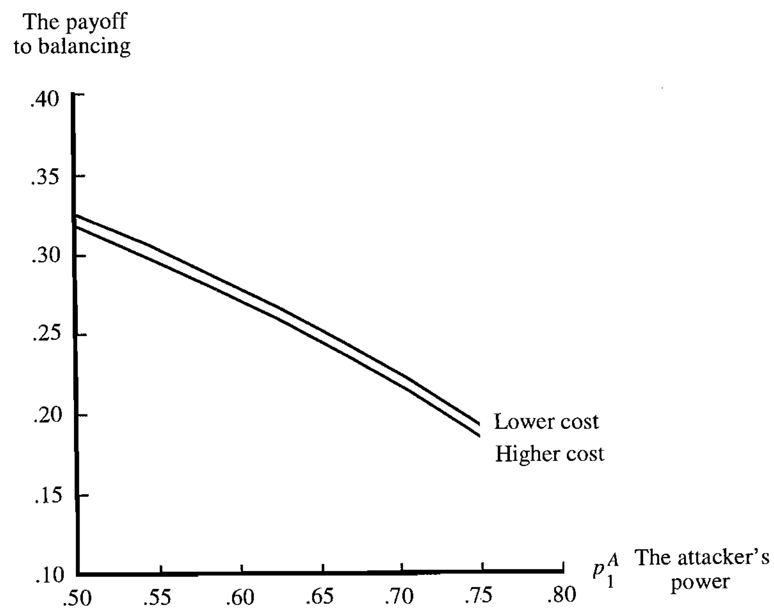
The effects of changes in the cost of fighting and of the degree to which military forces cumulate are depicted in figure 5.3. If S_1 balances by fighting with S_2 against A, then S_1 is certain to pay the cost of fighting. Accordingly, an increase in the cost of fighting should decrease the payoff to balancing, and this turns out to be the case. Figure 5.3a shows the effect of an increase in the cost.

The extent to which two states' military forces cumulate when they align with each other affects the probability that S_1 and S_2 will defeat the attacker. The more that these states' forces cumulate, the higher the probability of defeating A and the larger S_1's payoff to aligning with S_2. Figure 5.3b illustrates this in the case in which there are decreasing returns to scale ($g < 1$), constant returns to scale ($g = 1$), and increasing returns to scale ($g > 1$). As expected, the payoff to balancing increases in g.

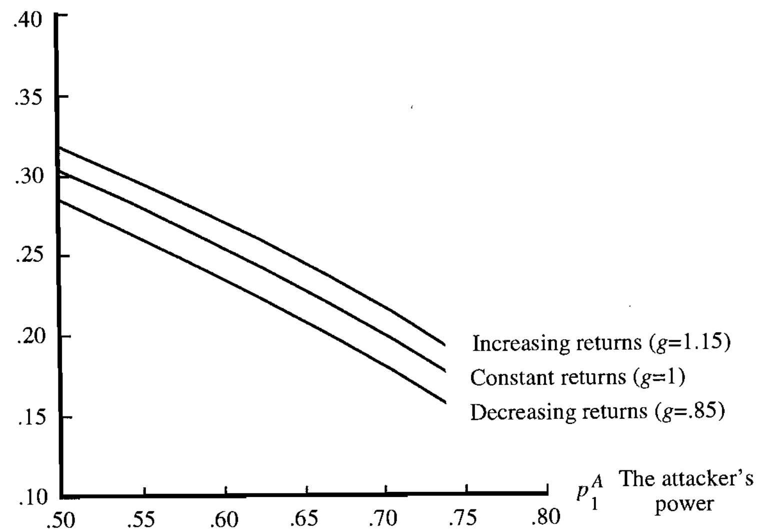
The payoff to waiting is somewhat more complicated because it incorporates several factors. Suppose A attacks S_2 and S_1 waits. Waiting ensures that S_1 makes it into the bargaining stage of the game. Therefore, S_1's payoff to waiting is determined by the outcome of the bargaining that takes place between S_1 and the victor in the conflict between A and S_2. As shown in chapter 3, this outcome depends in turn on whether the victor is satisfied or dissatisfied with S_1.

Consider first the case in which the victor is satisfied with S_1. That is, the victor is unwilling to use force against S_1 given the distributions of power and benefits that exist between the victor and S_1 at the beginning of the bargaining. Because the victor is unwilling to use force, S_1 does not have to make any concessions to appease its adversary and, therefore, maintains control over all of its territory. Thus, S_1's payoff to waiting in these circumstances is its payoff to controlling its original share of the territory.

If, by contrast, the victor is dissatisfied with S_1, then S_1 has to concede some of its territory to its adversary. And, the more powerful its adversary, the more that S_1 has to surrender. (Recall that S_1 must give its adversary control over an amount of territory equal in value to its adversary's payoff to fighting, and, therefore, the more powerful the adversary, the more S_1 must concede.)



(a) The effects of changes in the cost of fighting.



(b) The effects of changes in the returns to scale.

Figure 5.3 The effects of changes in the technology of coercion on the payoff to balancing

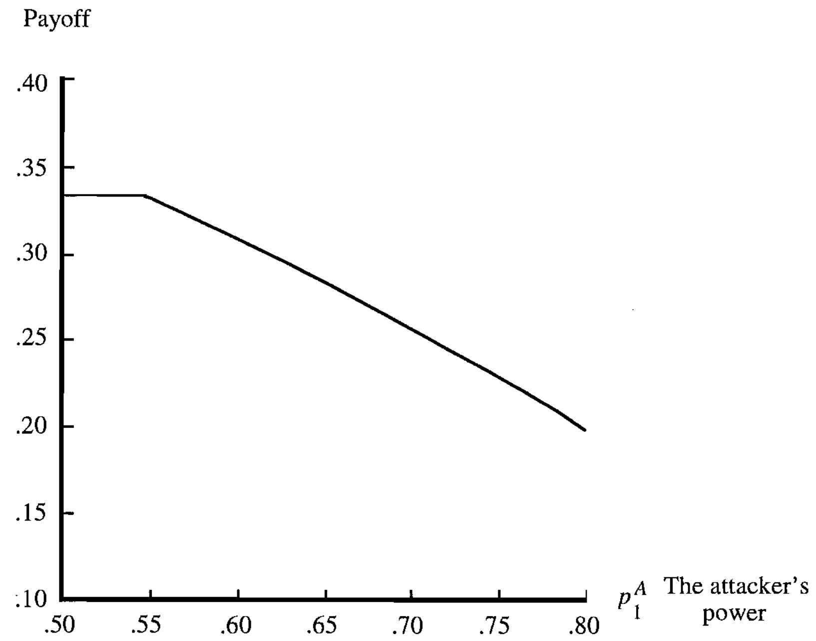


Figure 5.4 The payoff to waiting

The shape of S_1's payoff to waiting in figure 5.4 reflects these factors. As before, the graph begins at $p_1^A = \frac{1}{2}$ where the distribution of power mirrors the distribution of benefits. Then, as the potential attacker's capabilities begin to grow, p_1^A becomes a little larger than $\frac{1}{2}$. In these circumstances, the victor of the contest between A and S_2 is still satisfied with S_1, and, consequently, S_1 retains all of its territory and its payoff is $\frac{1}{3}$.

The reason why the victor is satisfied is that its military capability depends on two key factors: the initial capabilities of A and S_2, which the victor combines after defeating its adversary, and the extent to which these capabilities cumulate when combined. If the potential attacker is only a little stronger than it would be if the initial distribution of power among the three states exactly mirrored the distribution of benefits (i.e., if p_1^A is not too much larger than $\frac{1}{2}$) and if military capabilities do not cumulate too much (i.e., if g is not too big), then the victor's power grows in roughly the same proportion as its territory does when it incorporates the defeated state's military capabilities *and* its territory. This implies that the distribution of power between the victor and S_1

is approximately the same as the distribution of benefits between them. In these circumstances, the victor is satisfied and S_1 does not have to make any concessions.

If, by contrast, the attacker's military capabilities are relatively large so that p_1^A is large in comparison to A's territorial share of $\frac{1}{3}$, then when the victor in the war between A and S_2 combines both states' military capabilities, its power will be proportionately larger than its territory. In these circumstances, the victor is dissatisfied with S_1 at the beginning of the bargaining stage, and S_1 has to appease its adversary. This leaves S_1 with a payoff of less than $\frac{1}{3}$. Indeed, the stronger A, the stronger the victor of the struggle between A and S_2 is after combining the states' capabilities, and the more S_1 has to concede during the subsequent bargaining. Accordingly, S_1's payoff to waiting declines as p_1^A increases as figure 5.4 shows.[16]

A change in the technology of coercion, which increases the extent to which military forces cumulate, makes the victor in the fight between A and S_2 stronger in the bargaining with S_1. This has two effects on S_1's payoff to waiting. First, it shrinks the range over which S_1 can stand aside without then having to make concessions to the victor during the bargaining stage. That is, if military capabilities cumulate more (a higher g), the victor becomes dissatisfied with S_1 when A's military capabilities are smaller and p_1^A is lower. Accordingly, S_1's payoff to waiting begins to decline earlier, as it has to make concessions to satisfy its adversary at lower values of p_1^A. Second, if S_1 already had to make a concession to appease a dissatisfied victor, then S_1 would have to make an even larger concession if there were larger returns to scale. Thus, larger returns to scale reduce S_1's payoff to waiting. These effects are depicted in figure 5.5.

[16] These complications did not arise when calculating the payoff to balancing because the attacker's strength affects the chances that the balancing coalition of S_1 and S_2 prevails, but it does not affect whether or not there is a dissatisfied state in the bargaining round. Indeed, S_1 and S_2 are always satisfied with each other if their coalition prevails. Thus, S_1's payoff to balancing against A continually decreases as A becomes stronger.

The fact that S_1 and S_2 remain satisfied with each other regardless of A's strength follows from three assumptions. First, the initial distribution of power between these states mirrors the distribution of benefits between them ($p_2^1 = \frac{1}{2}$, $q_1 = q_2 = \frac{1}{3}$), and, therefore, these two states are initially satisfied with each other. Second, the distribution of power between S_1 and S_2 is unaffected by the defeat of A, and, third, the fighting that defeats A results in an initial division of A's territory between S_1 and S_2, which is proportional to the distribution of power between them. Consequently, the distributions of power and benefits between S_1 and S_2 mirror each other at the beginning of the bargaining stage.

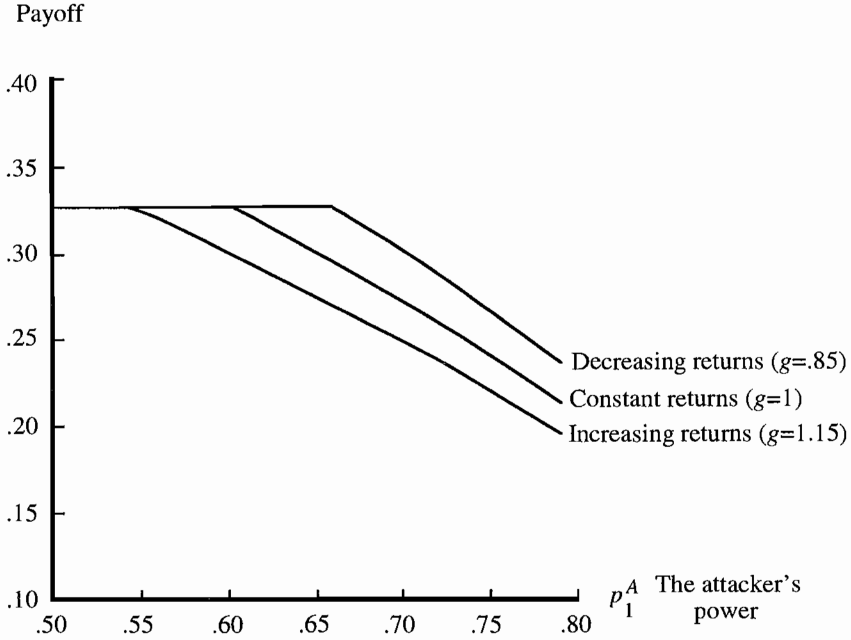


Figure 5.5 Returns to scale and the payoff to waiting

Finally, consider S_1's payoff to bandwagoning with A by joining in the attack on S_2. Here, S_1's payoff depends on what happens in the bargaining stage that takes place if the coalition of A and S_1 defeats S_2. If A is satisfied with S_1 following the elimination of S_2, then S_1 keeps all of its own territory plus the share of territory it takes from S_2. If, by contrast, A is dissatisfied with S_1, then S_1 has to appease A during the bargaining stage by conceding some territory. And, the stronger A, the more S_1 concedes and the lower is its payoff to bandwagoning.

These features are reflected in figure 5.6. As in the past, the distribution of power among the three states mirrors the distribution of benefits at $p_1^A = \frac{1}{2}$, given that each state initially controls one-third of the territory. If the potential attacker is only a little more powerful than this, then A would be satisfied with S_1 following the defeat of S_2. The reason is that because A and S_1 incorporate S_2's territory in proportion to the distribution of power between them, the distributions of power and territory at the start of the bargaining stage will roughly mirror each other if p_1^A is not too much larger than $\frac{1}{2}$.

As long as A would be satisfied with S_1 following the defeat of S_2 (i.e., as long as p_1^A is not too much larger than $\frac{1}{2}$), we might expect S_1 to

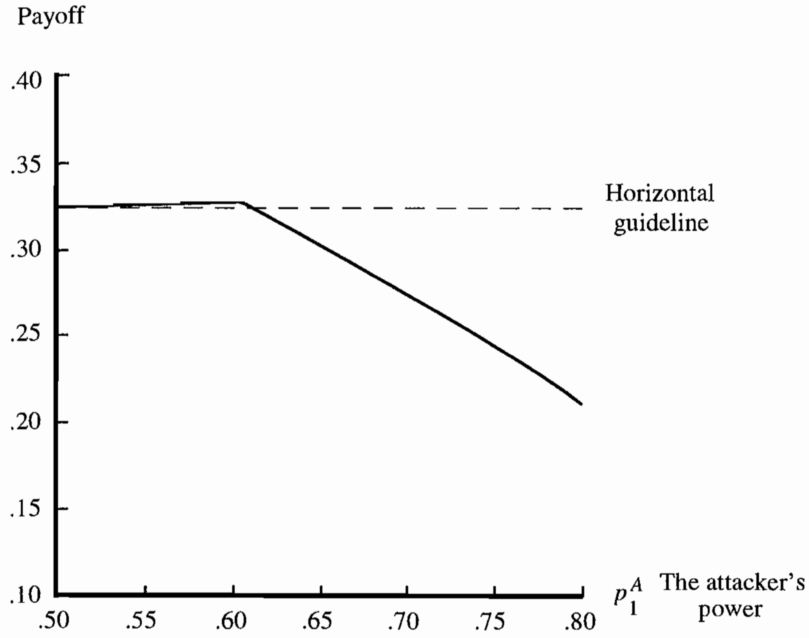


Figure 5.6 The payoff to bandwagoning

benefit from having a stronger coalition partner, and this is illustrated in figure 5.6. S_1's payoff to bandwagoning with A increases as A becomes stronger and p_1^A increases as long as p_1^A is not too much larger than $\frac{1}{2}$. (The horizontal guideline makes this easier to see.)

But this changes once the attacker is so strong that it would be dissatisfied with S_1 following S_2's elimination. S_1 no longer benefits from having a stronger coalition partner. Indeed, the stronger A is, the more S_1 has to concede and the lower its payoff. Accordingly, S_1's payoff to bandwagoning eventually begins to decline as p_1^A continues to rise.[17]

The effect of a change in the technology of coercion that increases the extent to which military capabilities cumulate makes the coalition of A and S_1 stronger. This in turn raises the coalition's probability of prevailing, but it does not affect the distributions of power and benefits

[17] Even when S_1 is not facing a dissatisfied coalition partner, an increase in its partner's strength has two opposing effects. It makes the coalition more likely to win, but it also reduces S_1's share of the spoils because these are divided according to the distribution of power between the states in the coalition. In the present formulation, the first effect dominates the latter, and S_1 benefits from having a stronger coalition partner.

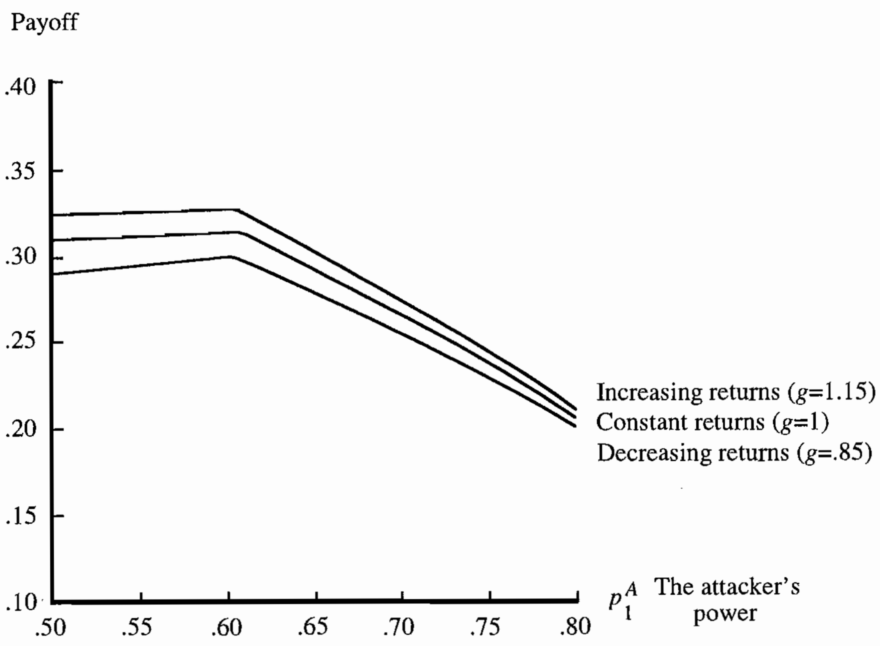


Figure 5.7 Returns to scale and the payoff to bandwagoning

between the coalition partners. Thus, the payoff to bandwagoning rises as the returns to scale increase as is shown in figure 5.7.

In sum, an increase in A's military capabilities may affect another state's payoffs to balancing, bandwagoning, and waiting through two channels. First, it may affect the likelihood that a state survives the first stage of the game and makes it into the bargaining round. Second, an increase in A's capabilities may also affect the distributions of power and benefits that exist between the two surviving states at the beginning of the bargaining. These distributions subsequently determine whether or not the territorial status quo changes during the bargaining. As we have seen, the first factor largely determines the payoff to balancing, both factors shape the payoff to bandwagoning, and the second factor decides the payoff to waiting.

The Alignment Decision

States do not balance, bandwagon, or wait because they attach intrinsic value to any of these behaviors. States align for instrumental reasons. Balancing, bandwagoning, and waiting are means through which a state

tries to obtain more basic ends. The questions to ask therefore are: What kinds of alignment decisions do the strategic incentives in the international system induce? Are there dominant patterns of state behavior? Do these incentives generally lead to balancing, and, if so, do states usually align against the most powerful state as Waltz (1979) and others argue or against the most threatening state as Walt (1987, 1988) claims? Or is bandwagoning more prevalent than balancing as Schroeder (1994a) believes? And, how do changes in the technology of coercion affect these decisions?

This section examines the states' alignment decisions in the model when the states are all equally willing to use force, i.e., when the states' marginal costs c_A, c_1, c_2 are the same. (This assumption will be relaxed below.) Balancing behavior is largely absent in these circumstances; states wait or bandwagon.

Whether a state waits or aligns with another depends on the extent to which there are increasing returns to scale in the aggregation of military capabilities. If capabilities do not cumulate very much, a state is more likely to "pass the buck" (Christensen and Snyder 1990) by letting others bear the burden and pay the cost of fighting. If, by contrast, there are larger returns to scale, then a state suffers a large adverse shift in the distribution of power if it waits while others fight. Worse, the victor's power grows disproportionately large compared to the additional territory it acquires from the defeated state. Thus, the victor will be dissatisfied with the state that waited, and the latter will have to make a large concession to appease the victor. This large concession makes waiting very costly if there are large returns to scale, and a state will join the fray by either balancing or bandwagoning to prevent the distribution of power from shifting against it. In this sense, the states are chain-ganged (Christensen and Snyder 1990) together.

A state faces a different trade-off when choosing between balancing and bandwagoning. *By aligning with the weaker of two other states, a third state makes itself more powerful with respect to its coalition partner than it would have been had it aligned with the stronger of the two other states.* Because of this more favorable distribution of power, a state is less vulnerable to being exploited by its coalition partner and better able to further its interests should the coalition prevail. More formally, a state will be more powerful in the bargaining stage of the model if it aligns with the weaker of the two other states and, consequently, will obtain a "better deal" from its coalition partner than it would have received from aligning with the stronger state. This consideration makes balancing more attractive than bandwagoning.

But the accounting mechanism of the model shows that there is an opposing factor at work. The advantages of the more favorable distribution of power created by aligning with the weaker of two possible coalition partners will only be realized if the coalition prevails. As Waltz puts it, "Secondary states, if they are free to choose, flock to the weaker side; for it is the stronger side that threatens them. On the weaker side, they are both more appreciated and safer, *provided, of course, that the coalition they join achieves enough defensive or deterrent strength to dissuade adversaries from attacking*" (1979, 127, emphasis added). But states are rarely sure of defeating or deterring another state and usually can only affect the chances of doing so. *A state is less likely to be part of a winning coalition if it aligns with the weaker side and more likely to be part of a winning coalition if it aligns with the stronger side*. This consideration makes bandwagoning more appealing.

Thus, the states face a complicated trade-off when deciding between balancing and bandwagoning. Balancing entails a lower probability of being part of a winning coalition but being relatively more powerful if the coalition prevails. Bandwagoning entails a higher probability of being part of a winning coalition but being relatively weaker if the coalition prevails. The latter usually outweighs the former in the model, and states generally bandwagon if there are large returns to scale.

To examine the alignment decisions in more detail, consider first the situation in which the states, in addition to being equally resolute, are all the same size and there are moderately large increasing returns to scale. In symbols, $q_A = q_1 = q_2 = \frac{1}{3}$ and $g = 1.15$. Suppose further that A attacks S_2. Then figure 5.8 plots S_1's payoffs to waiting, balancing against A by aligning with S_2, and bandwagoning with A by joining in the attack on S_2.[18]

The graph begins where the distribution of power among the three states exactly mirrors the distribution of benefits, i.e., at $p_1^A = q_A/(q_A + q_1) = \frac{1}{2}$. At this point, S_1's payoff to balancing equals its payoff to bandwagoning because A and S_2 are equal in size and power. Thus, it makes no difference to S_1 whether it aligns with S_2 to fight A or with A to fight S_2. However, S_1 has to bear the cost of fighting if it balances or bandwagons. If, by contrast, S_1 waits, it avoids these costs. Furthermore,

[18]These are also S_2's payoffs if A attacks S_1 because S_1 and S_2 are identical in size, power, and willingness to use force.

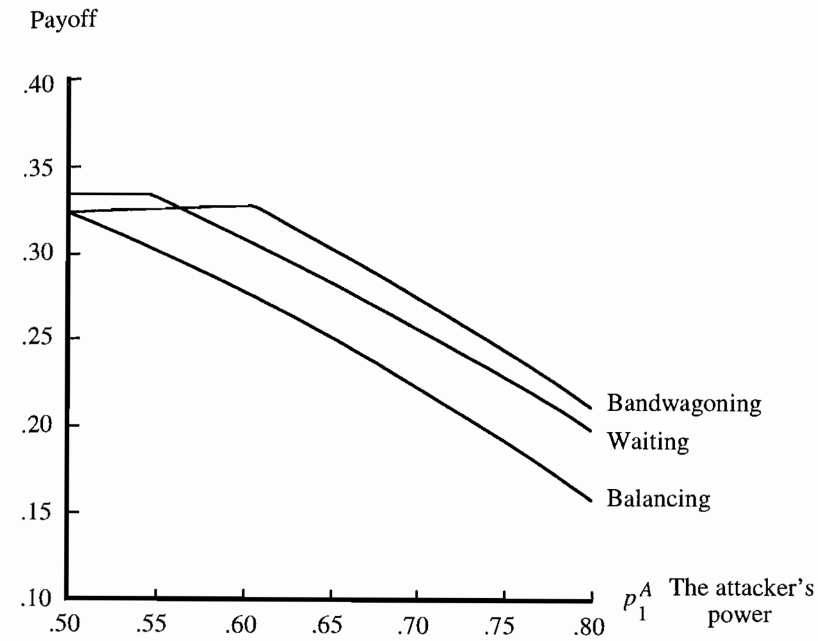


Figure 5.8 The payoffs to waiting, bandwagoning, and balancing

S_1 will not have to make any concessions during the bargaining because the distributions of power and benefits mirror each other. These factors make the payoff to waiting higher than the payoffs to bandwagoning or balancing. Consequently, S_1 prefers not to fight—even if aligned with another state—when the distribution of power reflects the distribution of benefits and fighting is costly. As was the case in the two-actor model in chapter 3, states are generally unwilling to use force to overturn the status quo when the distribution of benefits mirrors the distribution of power.[19]

Two aspects of figure 5.8 are especially noteworthy. First, S_1 never balances regardless of how strong A is. This contradicts the expectations of both the balance-of-power and balance-of-threat analyses. The former argues that S_1 should align against A because A is the stronger power.[20]

[19]This unwillingness depends in a multipolar setting on the extent to which military capabilities cumulate. If there are very large returns to scale, then the expected payoff to fighting as a member of a coalition may outweigh the costs even if the distributions of power and benefits among the states are identical.

[20]If $p_1^A > \frac{1}{2}$, then the probability that A will defeat S_2 is greater than $\frac{1}{2}$ since $p_1^A = p_2^A$, and therefore A is more powerful than S_2.

The latter also suggests that S_1 should align against A because A as the attacker is the source of the threat to the status quo. Yet, S_1 never aligns against A. It either waits or bandwagons.

The second important feature to note is the range of values of p_1^A over which S_1 waits and bandwagons. S_1 waits if A is only somewhat more powerful than it is when the distributions of power and benefits exactly mirror each other. (That is, S_1 waits if p_1^A is not much larger than $\frac{1}{2}$). But S_1 bandwagons if A is substantially more powerful. To see why, suppose p_1^A starts at $\frac{1}{2}$ and then increases. As just noted, S_1 prefers waiting at this point because fighting is costly. As p_1^A begins to increase, S_1's payoff to waiting initially remains constant because its adversary will still be satisfied during the bargaining stage, and, therefore, S_1 will not have to make any concessions. By contrast, S_1's payoff to bandwagoning with A initially increases because S_1 does better with a stronger coalition partner as long as that partner is not so strong that it will be dissatisfied with S_1 in the bargaining stage.

As A's military capabilities continue to rise, p_1^A continues to increase. Eventually, the victor in the struggle between A and S_2 will be so powerful after it has incorporated the defeated state's military capabilities that the victorious state will be dissatisfied with S_1 when the bargaining begins. If S_1 waits in these circumstances, it will have to appease its adversary during the bargaining; and the stronger the adversary, the more S_1 will have to concede. Accordingly, S_1's payoff to waiting begins to decline and continues to do so as p_1^A increases. At some point, waiting means such a large adverse shift in the distribution of power against S_1 that the cost of appeasing S_1's adversary during the subsequent bargaining becomes larger than S_1's cost to fighting in a coalition with A. At that point, S_1 switches from waiting to bandwagoning to forestall this adverse shift in the distribution of power.

We can think of this switch as a change from what Christensen and Snyder (1990) call buck-passing to chain-ganging. States pass the buck when they let others bear the costs of fighting and are chain-ganged together when no state can sit out a conflict because doing so would allow the distribution of power to shift too far against it. And, as we have just seen, S_1 changes from waiting to bandwagoning to avoid just such a shift.[21]

[21] Christensen and Snyder (1990) discuss buck-passing and chain-ganging solely in the context of the internal dynamics of alliances, whereas these terms are used here to refer to states' alignment decisions.

How do changes in the technology of coercion affect when S_1 switches from waiting to bandwagoning? Roughly, the more military capabilities cumulate, the more states tend to be chain-ganged together and the earlier they begin to bandwagon. To see this, observe that the larger the returns to scale, the more quickly S_1 faces a dissatisfied adversary if it waits and the more quickly its payoff to waiting begins to decline. If, by contrast, S_1 bandwagons, then larger returns to scale make the coalition of A and S_1 stronger and, thereby, increase S_1's payoff. The combined effect of these changes is that S_1 switches from waiting to bandwagoning sooner (i.e., at a lower p_1^A) as figure 5.9 indicates.

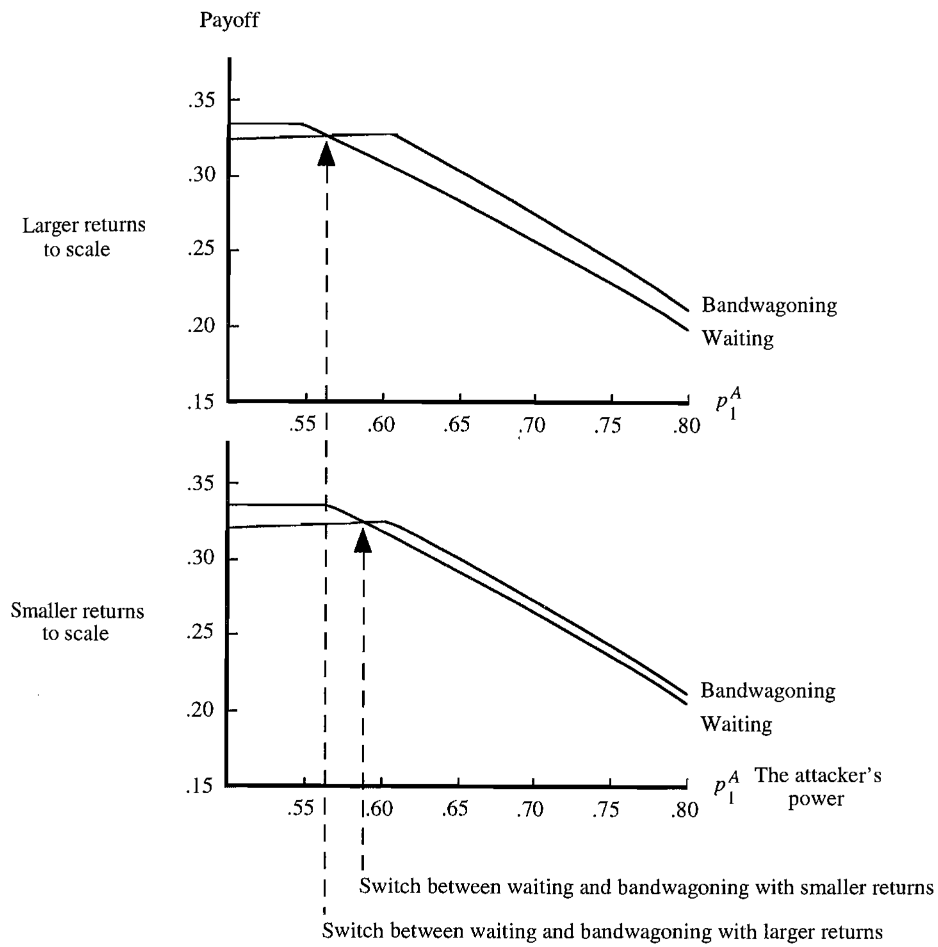


Figure 5.9 Shifts from waiting (buck-passing) to bandwagoning (chain-ganging)

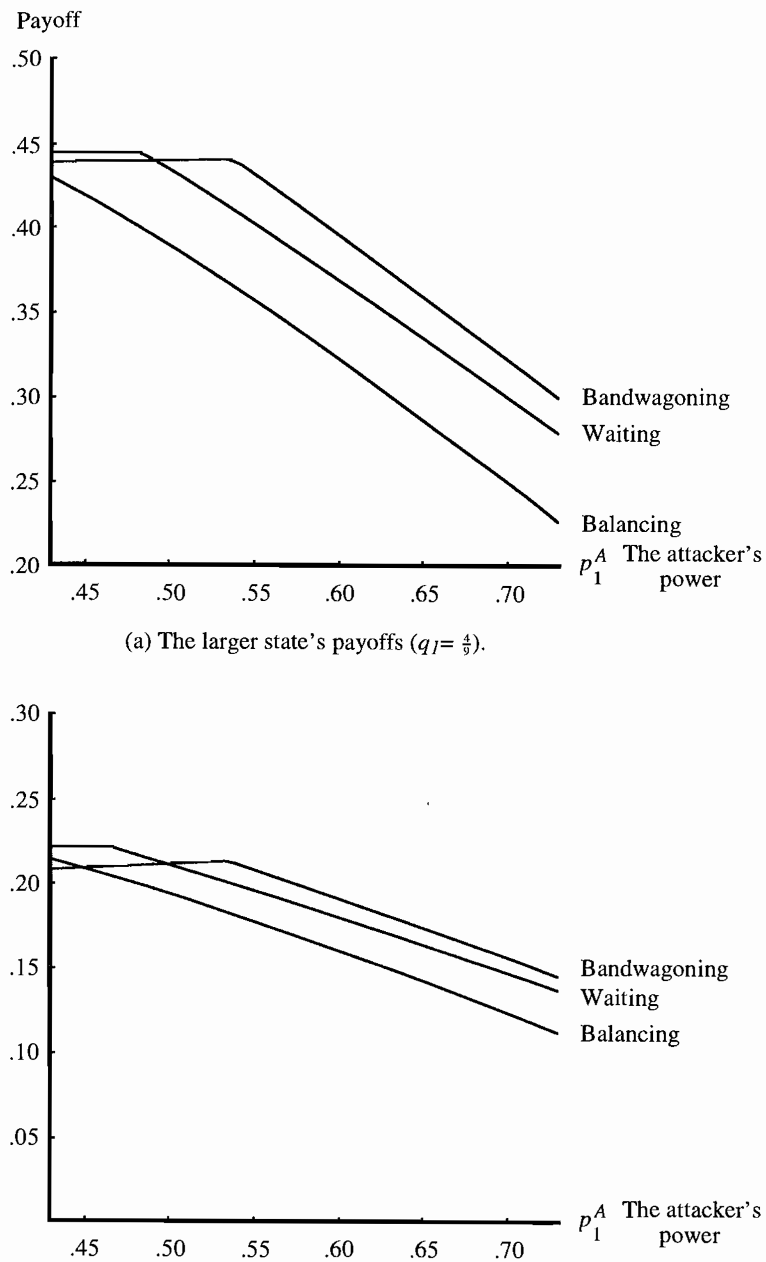
This section concludes with a discussion of two variants of the model. The preceding analysis has assumed that it does not matter how resources are combined. If one state conquers another, the victor is neither more nor less powerful than the coalition composed of these two states would have been. To relax this assumption, suppose that coalitions are relatively less efficient at combining resources. This change has no effect on a state's payoff to waiting but does lower the payoffs to balancing and bandwagoning. To wit, if a state waits while others fight, then no coalitions ever form and the fact that they would be less efficient if they had does not affect the payoff to standing aside. If, by contrast, a state bandwagons or balances, the coalition is less likely to prevail because it is relatively less efficient. This reduces the payoff to aligning and increases the range over which states wait.

The second variant examines the states' alignment decisions when the states differ in size. Suppose, for example, that S_1 is larger than the potential attacker and that S_2 is smaller. More concretely, assume that S_1 is, say, one-third larger than the potential attacker and S_2 is one-third smaller. In symbols, $q_A = \frac{1}{3}$, $q_1 = \frac{4}{9}$, and $q_2 = \frac{2}{9}$.

Figure 5.10 plots S_1's and S_2's payoffs starting from the point at which the distribution of power mirrors the distribution of benefits, i.e., $p_1^A = q_A/(q_A+q_1) = \frac{3}{7} \approx .43$. As in the past, the states prefer waiting to aligning. But the payoffs to balancing and bandwagoning are now different because the states are different sizes. In particular, the larger state S_1 prefers bandwagoning to balancing whereas the smaller state S_2 prefers balancing at $p_1^A \approx .43$. Or, put another way, if a state has to align with another state rather than wait, it prefers to align with the larger and, therefore, more powerful state. To wit, S_1 joins with A when choosing between A and the smaller S_2, and S_2 aligns with S_1 when choosing between A and the larger S_1.[22]

Increases in A's military capabilities have the same general effect on a state's alignment decision regardless of the state's size. As long as A is not too powerful, i.e., as long as p_1^A is not too large, the cost of appeasing a dissatisfied adversary during the bargaining phase is less than the cost of fighting, and the state waits. Once A becomes sufficiently powerful, waiting entails accepting too adverse a shift in the distribution

[22] Appendix 5 shows that if the distribution of power mirrors the distribution of benefits, then the conclusion that the payoff to aligning with the larger state is greater than the payoff to aligning with the smaller state holds in general and not just for the particular values of $q_A = \frac{1}{3}$, $q_1 = \frac{4}{9}$, and $q_2 = \frac{2}{9}$.



(a) The larger state's payoffs ($q_1 = \frac{4}{9}$).

(b) The smaller state's payoffs ($q_2 = \frac{2}{9}$).

Figure 5.10 The payoffs of differently sized states of power and subsequently having to concede too much to the other state during the bargaining. This makes waiting too costly, and the state enters the conflict by bandwagoning with the attacker in order to avert this shift.

The Decision to Attack

In deciding whether or not to attack, the potential attacker tries to anticipate the other states' subsequent alignment decisions and acts accordingly. This section traces the effects of those anticipations on the attacker's decision. As will be seen, the potential attacker does not attack if the distribution of power roughly reflects the distribution of benefits and the returns to scale are not too large. If, by contrast, there is a moderate disparity between the distributions of power and benefits, the potential attacker attacks the weaker of the other two states in the correct expectation that the stronger state will find fighting too costly and will wait. War in these circumstances remains confined to the attacker and its immediate target. If the attacker is still more powerful, no state can afford to wait and allow the distribution of power to shift significantly against it. In these circumstances, both S_1 and S_2 will bandwagon. Anticipating this reaction, A now attacks the larger of the two states, after which the smaller state aligns with A.

These decisions illustrate the credibility problem that haunts alliances. If S_1 and S_2 could commit themselves to balancing through, say, forming an alliance, then their combined capabilities would deter the potential attacker from actually attacking and this would make both S_1 and S_2 better-off. However, S_1 and S_2 cannot *commit* themselves in advance to balancing. And, as we have seen, these states prefer bandwagoning or waiting to balancing when they must actually decide what to do. This undermines the credibility of any previous promise to balance.

Figure 5.11 plots the potential attacker's payoffs given that the states are equally resolute and that S_1 is one-third larger than the potential attacker and that S_2 is one-third smaller than the potential attacker (i.e., the initial distribution of territory is $q_A = \frac{1}{3}$, $q_1 = \frac{4}{9}$, and $q_2 = \frac{2}{9}$). If the potential attacker waits, the territorial status quo goes unchallenged and A's payoff remains constant at $\frac{1}{3}$.

The payoff to attacking the smaller state S_2 depends on the potential attacker's military strength. At the outset where the distribution of power mirrors the distribution of benefits (i.e., at $p_1^A \approx .43$), A's payoff to striking S_2 is less than its payoff to waiting. This follows because the

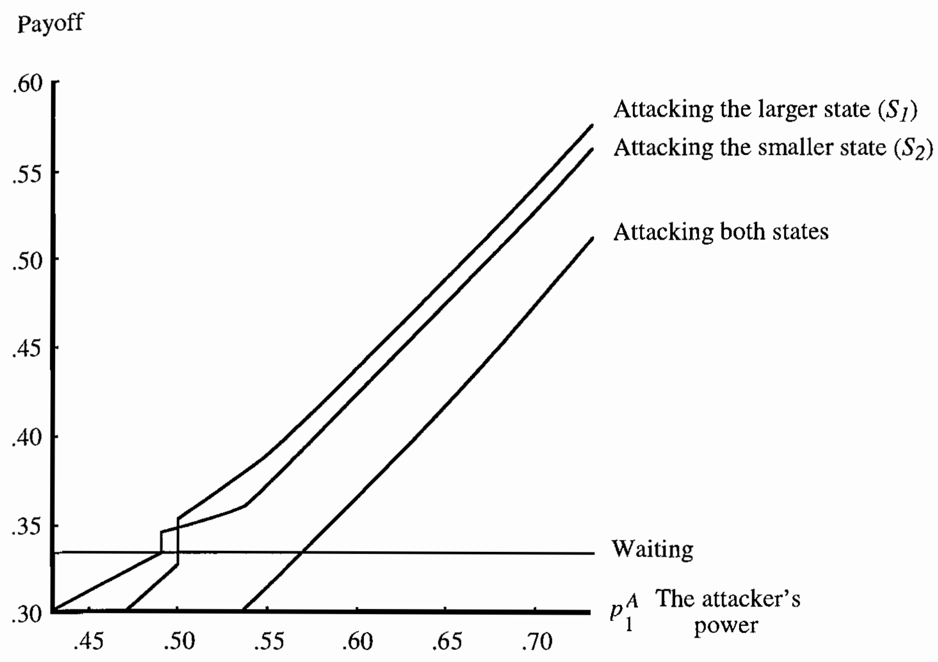


Figure 5.11 The potential attacker's payoffs

conflict between A and S_2 remains completely isolated and unaffected by S_1's presence. That is, S_1 waits while A and S_2 fight (see figure 5.10a) and then does not have to offer any concessions in the subsequent bargaining because the victor of the struggle between A and S_2 is satisfied with S_1. Since the conflict between A and S_2 remains isolated, then the analysis of the bilateral bargaining game in chapter 3 shows that A is satisfied with S_2 and therefore prefers not to attack S_2 as long as fighting is costly and the distribution of benefits between A and S_2 reflects the distribution of power between them.

As the potential attacker's capabilities start to rise, p_1^A begins to increase and A's payoff to attacking S_2 also grows. This higher payoff results solely from the fact that A is stronger and therefore more likely to defeat S_2 in a war that still remains isolated. To see that S_1's presence continues to have no effect, observe from figure 5.10a that S_1 still prefers to wait and does not have to make any concessions during the bargaining phase until $p_1^A \approx .48$ (which is where S_1's payoff to waiting begins to decline because it starts making concessions). Consequently,

S_1 continues to wait and subsequently refuses to make any concessions during the bargaining as long as $p_1^A < .48$.

As p_1^A increases a little more, A's payoff to attacking S_2 jumps up, and figure 5.10a explains why. S_1 switches at this point from waiting to bandwagoning. This change in S_1's strategy has two opposing effects on A's payoffs to attacking S_2. First, it significantly increases the military capabilities brought to bear against S_2, which boosts A's payoff by raising the chances that A will defeat S_2. But, second, the spoils of victory now have to be divided with A's new coalition partner S_1, and this tends to reduce A's payoff. The former, however, dominates the latter in the model, and the net effect on A's payoff is an upward jump.

As the potential attacker becomes still stronger, p_1^A continues to increase and A's payoff to attacking S_2 rises moderately until about $p_1^A \approx .53$. This gradual rise is solely due to the higher probability that the coalition of A and S_1 prevails. If it does, then there will be no further changes in the status quo during the bargaining stage as neither A nor S_1 will be dissatisfied. However, once p_1^A exceeds .53, A will be dissatisfied with S_1 if their coalition defeats S_2. S_1 therefore begins to have to make concessions during the bargaining stage, and S_1's payoff to bandwagoning starts to decline (see figure 5.10a). A's overall payoff to attacking S_2 therefore rises more rapidly.

The potential attacker's payoff to striking the larger state S_1 resembles its payoff to fighting the smaller state S_2. At the outset where the distributions of power and benefits are identical (i.e., at $p_1^A \approx .43$), A's payoff to attacking S_1 is less than its payoff to waiting as shown in figure 5.11. As long as A is not too powerful, S_2 waits (see figure 5.10b) and A's payoff to attacking S_1 rises as A becomes stronger and p_1^A increases. This changes at about $p_1^A \approx .50$, where waiting becomes too costly for S_2 and it enters the fray by aligning with A. At this point, A's payoff to attacking S_1 jumps up and thereafter rises as p_1^A increases.

Finally, figure 5.11 also depicts A's payoff to attacking both S_1 and S_2 at once. The simplifications made in the model imply that this payoff equals A's payoff to attacking one of the other states given that the third state subsequently balances. Increases in the attacker's capabilities then simply raise the probability that A would prevail over the de facto coalition of S_1 and S_2, and A's payoff rises smoothly in p_1^A.

Now consider what these payoffs imply about the potential attacker's decision. A prefers to wait and the status quo goes unchallenged when the distribution of power approximates the distribution of benefits. This result parallels that derived from the bargaining model in chapter 3.

As the potential attacker becomes more powerful, waiting remains optimal until $p_1^A \approx .49$. At that point, A begins to prefer attacking the smaller state S_2 in the correct expectation that S_1 will join in the attack. A continues to prefer attacking the smaller state S_2 until A is so powerful that S_2 could no longer afford to stand aside if A attacked S_1. Once the potential attacker becomes this strong, it attacks S_1 and thereby drags S_2 into the conflict on A's side.

Three important points about balancing emerge from this discussion of the potential attacker's decision. First, balancing is generally absent in the model when the states are equally resolute. States typically do not align against the most powerful state or against the most threatening state. But, second, there is a rough sense in which a balance of power might be said to occur: if A attacks, then for all but a small range of p_1^A, it attacks the larger of S_1 and S_2. This suggests that the two most powerful states, namely A and S_1, will usually be on opposite sides, and, in this loose way, balances of power might be said to form even though the states themselves do not balance by opposing the strongest or most threatening state.

Finally, note that if S_1 and S_2 could commit themselves to balancing through, say, an alliance, then both would be better-off as long as A is not too powerful. To see this, observe that A's payoff to fighting both S_1 and S_2 is less than its payoff to waiting as long as p_1^A is less than .57 or so (see figure 5.11). Consequently, A would not attack the alliance of S_1 and S_2 as long as p_1^A is less than .57 and *the alliance is credible*. The original status quo, therefore, would remain unchanged, and S_1 and S_2 would be better-off. However, the states cannot commit themselves to balancing, and the credibility of any promise to do so is undone by the fact that if a state ever had to follow through on its promise, it would do better by bandwagoning instead of balancing. A therefore attacks (as long as p_1^A is at least .49).

This reasoning suggests that states at times would like to be able to create a "commitment device." That is, they would like to find ways of making it costly for them to renege on their promises to support each other. Such steps make an alliance more than a scrap of paper; they raise the cost of waiting or bandwagoning. This in turn makes balancing more likely and the promise to do so more credible. Indeed, creating a structure that made it costly for the United States not to support Western Europe in the event of a Soviet invasion was one of the guiding principles behind the North Atlantic Treaty Organization (NATO). The remark generally attributed to Lord Ismay, NATO's first secretary general, aptly captures this concern (as well as two others): the purpose of NATO is "to keep the Russians out, the Americans in, and the Germans down."[23]

Alignment Decisions with Asymmetric Resolve

Until now the analysis has focused on the symmetric case in which all of the states were assumed to be equally resolute. This is the simplest case and the one most in keeping with a "billiard-ball" approach to international relations theory in which states are presumed to be identical. But as we have seen, balancing is generally absent in the model if states are equally resolute, and this seems to conflict with the historical record. Whether or not balancing is historically more prevalent than bandwagoning, states do balance at times.

This conflict between the formal implications of the model and historical experience suggests that the circumstances that give rise to balancing are more restricted than has previously been appreciated. The model shows, for example, that the absence of a central authority in a strategic setting in which states can use power against each other does not in and of itself induce balancing: There is no central authority in the model and the states can use power to further their ends, yet balancing is absent. The model therefore provides a counter-example to the claim that anarchy implies a strong tendency to balance.[24] This counter-example in turn indicates that our existing understanding of the circumstances that produce balancing is at best incomplete.

But, the conflict between the alignment game and historical experience can do more than show us that we "know" less than we thought we did when we impose the accounting standards of a formal model. This "evolution of ignorance" (Krugman 1995, 79) may also motivate us to take the next step in the modeling dialogue. In particular, it may prompt us to ask, what might account for balancing that is not currently in the model?

The previous discussion of bargaining points to a possible answer. Actions generally reveal information about an actor. More conciliatory offers, for example, suggest a state is less willing to use force, whereas

[23] If alliances are costly to break, then states can use them as a costly signal of their resolve to support each other. Fearon (1994), Morrow (1994), Nalebuff (1991), and Smith (1998) develop this costly signaling perspective formally.

[24] For an example of the claim that anarchy induces balancing, see Waltz (1979, 114–28, especially 121).

an attack demonstrates a higher level of resolve. Germany's calculations on the eve of the First World War illustrate the point.

Germany believed that if Russia mobilized first and therefore appeared to be the aggressor, then Britain would remain neutral and not align with Russia against Germany. Consequently, German Chancellor Bethmann Hollweg delayed Germany's general mobilization even in the face of a partial Russia mobilization in the hope that this delay would force Russia to be the first to order a general mobilization. According to the minutes of a meeting with the chief of the general staff on the eve of war, German Chancellor Theobald von Bethmann Hollweg argued:

> We must wait for this [i.e., Russia's general mobilization] because otherwise we shall not have public opinion on our side either here or in England. The second was desirable because in the Chancellor's opinion *England* could not stand by Russia if Russia unleashed the furies of general war by attacking Austria and so assumed responsibility for the great mess.[25]

In other words, whether Germany or Russia attacked first would reveal information about their aggressiveness and this revelation would affect Britain's alignment decision.[26]

This section draws on the idea that an attack may indicate that the attacker is more willing to use force (i.e., has a lower marginal cost) than the other states. In particular, the section examines how states align when the attacker is much more willing to use force than the other states. Do bandwagoning and waiting continue to be the predominant behaviors, or does the aggressiveness of the attacker induce states to begin to balance by aligning against the attacker?

We might expect intuitively that balancing is more likely if the potential attacker is more willing to use force than the other states and if there are large returns to scale. The latter condition means that a state suffers a large adverse shift in the distribution of power if it waits. This makes waiting very costly and induces states to align. But, as we have seen, this

[25] Quoted in Fischer (1975, 494). Still hoping that Britian might remain neutral, Germany subsequently fabricated French border violations in order to justify Germany's invasion of Belgium. In the end, of course, all of these international efforts failed as Britain entered the war against Germany. But the attempt to paint Russia as the aggressor was more successful domestically in that Bethmann Hollweg did secure the support of the Social Democrats in the Parliament (Fischer 1975, 470–515). See Fischer (1988) and Levy (1990/91, 163–70) for further discussions of Germany's beliefs about British neutrality.

[26] And, of course, believing that Britain would be influenced by who mobilized first, Germany tried to use this strategically to manipulate British behavior.

decision to align generally leads to bandwagoning rather than balancing if all of the states are equally willing to use force.

But the terms of the trade-off between balancing and bandwagoning change if the attacker is more willing to use force than the other two states. If a state bandwagons with an attacker that is more willing to use force, then the bandwagoning state has to make larger concessions to the attacker during the bargaining stage. This lowers the payoff to bandwagoning and makes balancing more likely.

To trace this intuition more formally, suppose the potential attacker is more willing to use force than the other states (i.e., c_A is smaller than c_1 and c_2) and consider the effects of this change on the states' payoffs. The expected payoff to waiting decreases, because waiting implies having to make larger concessions during the bargaining stage if the attacker prevails during the first stage. The payoff to bandwagoning also declines for the same reason. If the attacker is more aggressive, the bandwagoning state has to offer more during the bargaining phase.

By contrast, the payoff to balancing remains the same even if the attacker is more willing to use force. Regardless of the attacker's cost, the balancing coalition either eliminates the attacker or is eliminated by it. If the coalition prevails, the attacker's cost does not affect the states' payoffs because the attacker's cost has no effect on the subsequent bargaining between the coalition partners. And, of course, the attacker's cost does not have any effect on the states' payoffs if they are eliminated. Thus, the net effect of the attacker's being more willing to use force reduces the payoffs to waiting and bandwagoning while leaving the payoff to balancing unchanged. This makes balancing more likely.

In brief, the payoffs to waiting and bandwagoning decline relative to balancing when the attacker is more aggressive than the other states. Figure 5.12 illustrates these effects when there are large returns to scale (say $g = 1.25$) and all of the states are equal in size.[27] At $p_1^A = \frac{1}{2}$, the distribution of benefits mirrors the distribution of power, and S_1's payoffs to balancing and bandwagoning are equal because A and S_2 are equivalent coalition partners (i.e., A and S_2 are the same size, equally powerful, and S_1 would not have to make a concession to either A or S_2 during any subsequent bargaining). The payoffs to bandwagoning and balancing are also greater than the payoff to waiting because the returns to scale are large. Larger returns to scale mean that if S_1 waits, the victor in the conflict between A and S_2 will be stronger and this implies that S_1 will have to make larger concessions during the bargaining phase. Waiting, therefore, entails accepting a large, adverse shift in the distribution of power, and to forestall this possibility, S_1 would enter the fray by either balancing or bandwagoning.[28]

As the potential attacker's capabilities begin to grow, A becomes more powerful and p_1^A increases. This increase reduces S_1's payoffs to balancing against A. S_1's payoff to waiting also goes down as A becomes stronger, because S_1 will have to concede more during the bargaining phase. By contrast, S_1's payoff to bandwagoning by aligning with A initially increases because S_1 benefits from having a stronger but still satisfied ally. Thus, the difference between the payoffs to bandwag-

[27] In the numerical illustrations, S_1's and S_2's costs are assumed to be $c_1 = c_2 = .15$ as before, but $c_A = .005$.

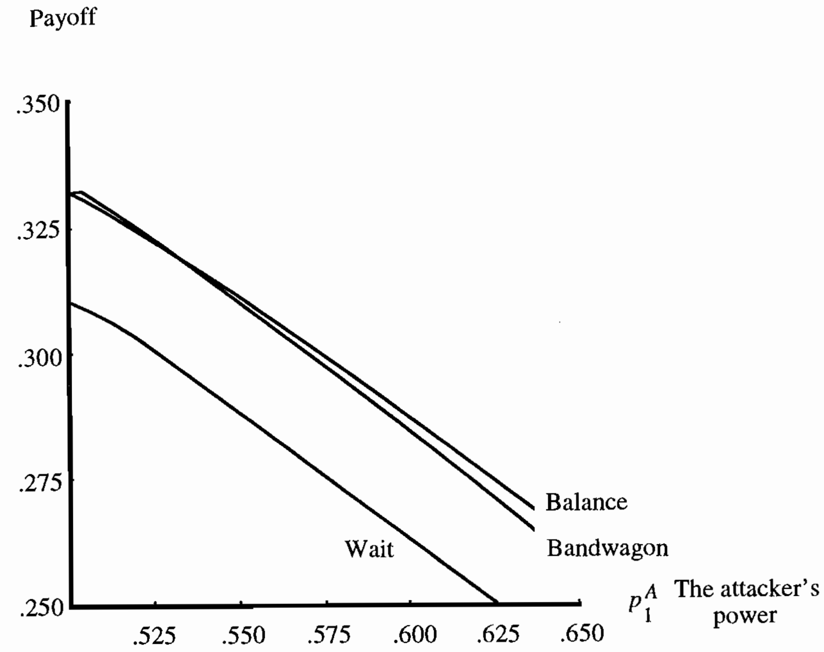


Figure 5.12 The payoff when the attacker is resolute and there are large returns to scale

[28] This contrasts with the situation depicted in figure 5.8 where the returns to scale are smaller. In that situation, S_1 did not have to make any concessions if it waited at $p_1^A = \frac{1}{2}$, and the payoff to waiting was greater than the payoff to bandwagoning or balancing. S_1, however, still prefers peace to war in figure 5.12. Athough its payoffs to balancing and bandwagoning are greater than its payoff to waiting, the former are smaller than S_1's status quo payoff of $\frac{1}{3}$.

oning and balancing, which is plotted in figure 5.13a, initially grows. In these circumstances, S_1 prefers to bandwagon.

This desire to bandwagon soon changes as A becomes stronger. Because A is very willing to use force, it quickly becomes sufficiently strong that S_1 will have to make some concessions to A if S_1 bandwagons with A and they defeat S_2. This occurs at $p_1^A \approx .504$ where S_1's payoff to bandwagoning begins to decline in figure 5.12 and where the difference between the payoffs to bandwagoning and balancing peaks in figure 5.13a.

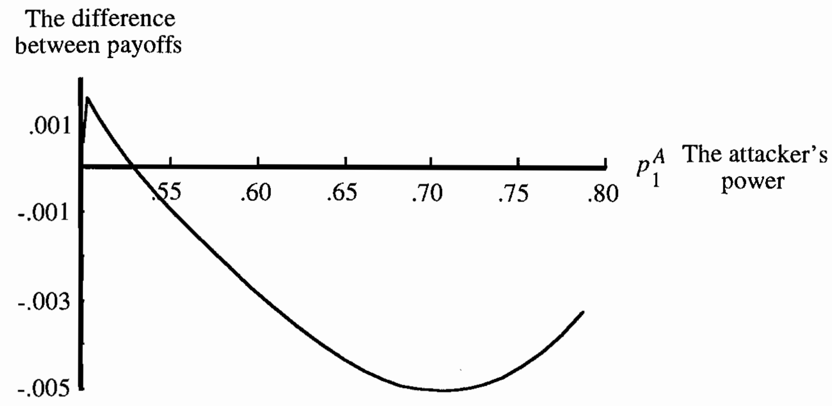
As A becomes still stronger, S_1's payoff to bandwagoning, like its payoff to waiting and balancing, decreases. Indeed, the gap between S_1's payoff to bandwagoning and balancing narrows. At $p_1^A \approx .54$, the cost of appeasing A in the bargaining stage starts to outweigh the benefits of being more likely to prevail in the first stage. The payoff to balancing exceeds the payoff to bandwagoning, and S_1 begins to balance.

Figure 5.13b depicts the potential attacker's payoffs to waiting and to attacking either S_1 or S_2.[29] Over the range $p_1^A = .5$ to $p_1^A \approx .54$, S_1 and S_2 bandwagon with A if A attacks the other state (see figure 5.13a). That is, S_1 and S_2 are chain-ganged to A. Thus, A's payoff to attacking rises as it becomes stronger and p_1^A increases over the range from .5 to .54. However, S_1 and S_2 switch from bandwagoning with A to balancing against it at $p_1^A \approx .54$, and A's payoff to attacking drops. Indeed, A's payoff to attacking drops below its payoff to waiting over the range .54 to .56 and A is deterred from attacking. Balancing over this range preserves the peace. However, once A becomes sufficiently powerful ($p_1^A > .56$), its payoff to fighting the coalition of S_1 and S_2 exceeds the payoff to waiting and A attacks.

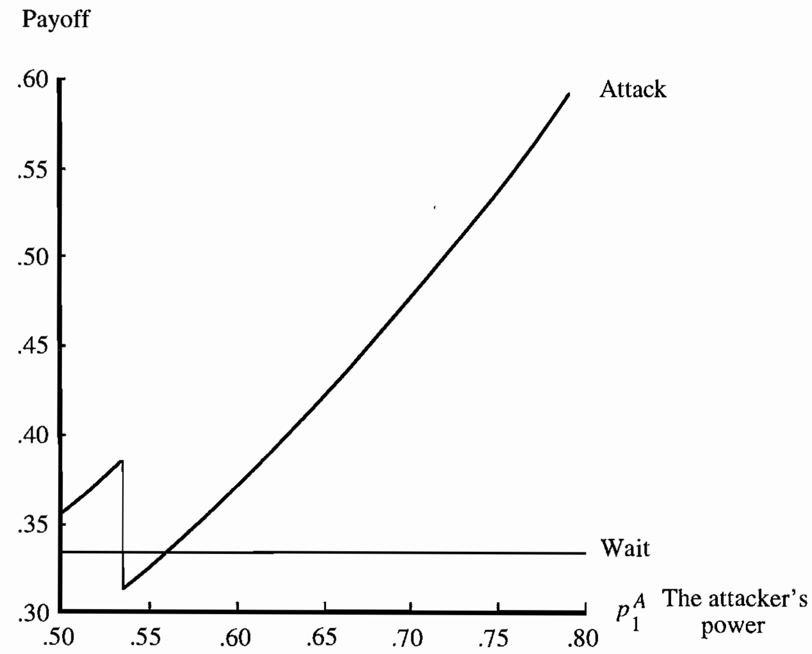
In sum, balancing seems to take place in a limited set of circumstances. If there are large returns to scale in the cumulation of military capabilities, then waiting implies accepting a large, adverse shift in the distribution of power and this makes it too costly to do. If the attacker is also much more willing to use force than the other states, then bandwagoning with the attacker means having to make large concessions to the attacker during any subsequent bargaining. This makes bandwagoning very costly, too, and the combination of large returns to scale and a very resolute attacker lead to balancing.

Pushing this idea beyond the present model, a state faces a difficult inference problem when it sees one state attack another. Does the at-

[29] A's payoffs to attacking S_1 or S_2 are equal because S_1 and S_2 are identical in size and territory and make the same alignment decisions.



(a) The payoff to bandwagoning less the payoff to balancing.



(b) The attacker's payoffs.

Figure 5.13 Balancing, bandwagoning, and attacking tack indicate that the attacker is generally aggressive and willing to use force or just that it is dissatisfied with the particular state it attacked? Does the attacker have limited or unlimited ambitions? (Putting these questions more formally, if A attacks S_2 should S_1 infer that A is also willing to use force against S_1 or just that A was dissatisfied with S_2 and that A is relatively unwilling to use force against S_1? Both of these interpretations are consistent with A's attack on S_2.) To the extent that an attack is interpreted as a sign that an attacker is generally aggressive, other states will tend to balance, whereas they will tend to wait if they read the attack as resulting from a particular dissatisfaction. This suggests that balances of power are more likely to form after a state has repeatedly demonstrated a willingness to use force, as did France under Louis XIV and Napoleon and Germany under Hitler. Louis XIV, for example, frustrated other states' attempts to stand aside or bandwagon "because France would not let their efforts succeed; they resisted [by balancing] because France kept on attacking them" (Schroeder 1994a, 135). Resistance to Hitler also solidified after he occupied the rump of Czechoslovakia and thereby demonstrated that his aims were not limited to incorporating Germans in the Reich.

Conclusion

The preceding analysis of the relation between the distribution of power and states' alignment decisions is exploratory and tentative. It is at most an early effort in the modeling enterprise. As such, it studies these decisions in the simplest formal setting. Yet even in this simple setting, states' alignment decisions reflect a complex and delicate trade-off of opposing influences. Figure 5.14 summarizes some of these decisions and the factors affecting them. Generally speaking, states tend to wait if the extent to which military capabilities cumulate is low. In these circumstances, the distribution of power does not shift significantly against a state if it waits and doing so avoids the cost of fighting. This situation is represented in the left column.

If there are large returns to scale (the right column), waiting brings a large adverse shift in the distribution of power. Standing aside while others fight thus means that the victor's power grows disproportionately more than the benefits it gains by defeating its adversary. The victor, therefore, is dissatisfied after defeating its first adversary and has to be appeased by the state that waited. The prospect of having to grant these

Attacker's Relative Resolve		Returns to Scale in the Aggregation of Military Capabilities: low	moderate	high
	equal	wait	wait or bandwagon	bandwagon
	high	wait	wait or bandwagon	balance

Figure 5.14 The alignment decision

large concessions makes waiting very costly and induces a state to join the fray by aligning with one of the other states.

Whether a state balances or bandwagons when it enters the conflict depends on how it resolves a trade-off. Balancing with the weaker state offers a lower probability of prevailing but a more favorable distribution of power if the coalition prevails. Bandwagoning, by contrast, offers a higher probability of prevailing but a less favorable distribution of power. The latter dominates the former if the states are equally resolute (the upper-right cell), and the states bandwagon. But, the former dominates the latter and states balance if the attacker is much more willing to use force than the other states, as in the lower-right cell (see figures 5.12 and 5.13 for an example).

If there are moderate returns to scale (the middle column), the states generally bandwagon or wait depending on the interaction of the size of the returns to scale and the attacker's willingness to use force. The less military capabilities cumulate, the less the distribution of power shifts against a state if it waits and the larger the benefit of avoiding the cost of fighting. A state is more likely to wait in these circumstances. However, the more willing the attacker is to use force compared to the other states, the more likely it is to be dissatisfied during the bargaining stage even if there has only been a relatively small shift of power. This makes waiting costly and bandwagoning more likely.

These results are consistent with some of the claims made about the relation between the distribution of power and states' alignment decisions and are inconsistent with others. By themselves, anarchy and the desire to survive do not produce a general tendency for states to balance against the most powerful state. The strategic environment in the model is anarchic, and the states, by trying to maximize the territory under their control, are necessarily also motivated to survive. Yet balancing is relatively rare in the model. Balances of power do sometimes form, but there is no general tendency toward this outcome. Nor do states generally balance against threats. Even when the attacker is much more willing to use force than the other state and in this sense the most threatening state, waiting and bandwagoning occur at least as much as balancing in the model.

The overall impression of the results is more akin to Schroeder's (1994a, 1994b, 1995) reading of European history. States frequently wait, bandwagon, or, much less often, balance. Balancing, when it does occur, is usually a response to a state—like France under Louis XIV or Napoleon, or Germany under Hitler—that has shown itself to be much more willing to use force than other states.

6

Conclusion

States do interact in a Hobbesian state of nature in which there is no superior authority to prevent one state from using the means of power to its own advantage and to the possible disadvantage of others. But this characterization of the states' strategic environment is too broad to be very useful. The strategic arena in which states interact can be described more narrowly and more productively in terms of the commitment issues, informational problems, and the underlying technology of coercion. These strategic features define the trade-offs states confront, and the previous chapters used a series of game-theoretic models to examine how states resolve those trade-offs.

This chapter does three things. First, it briefly reviews the results of this examination and compares the findings to existing arguments. The chapter then considers some of the models' limitations and offers a few conjectures about what will happen when some of these restrictive assumptions are relaxed in future rounds of the modeling dialogue. The chapter concludes with a discussion of the relation between international relations theory and political science more generally. International relations theory has often used the assumption of anarchy as a kind of analytic wall to set it off and separate it from the from the rest of political science. Exactly the opposite perspective emerges from the preceding chapters. Viewing the strategic environment confronting states in terms of commitment issues, informational problems, and the technology of coercion serves as a bridge to other areas in political science.

Some New and Old Results

One of the ways that a state can respond to a threat is by reallocating resources it already controls. However, internal balancing entails a trade-off. The more a state allocates toward satisfying current wants, the less it can devote to the means of military power and the weaker it will be in the future.

The key to resolving this guns-versus-butter trade-off turns out to be the expected payoff to attacking. If the payoff to attacking increases rel-

The second remark focuses on the finding that faster shifts are no more likely to break down in war than slower shifts. The overall probability of breakdown is the probability that R rejects an offer at some point in the shift in power. This is just $Z(\widetilde{r}_T)$. But $\widetilde{r}_T = \overline{p} - U_R(x_T^*)$, so the probability of breakdown is a function of D's final offer x_T^*. Expression A4.2 however, formally establishes that this offer is independent of the speed of the shift Δ.

Appendix 5

The Formalities of the Alignment

This appendix formally specifies the alignment game that was discussed informally in chapter 5. The appendix then establishes a claim made in that chapter: If the distribution of power mirrors the distribution of benefits and the potential attacker actually attacks one of the other states, then the third state's payoff to aligning with the larger of the other two states is greater than its payoff to aligning with smaller of the other two states (see page 182 above). More concretely, if A attacks S_1 when the distributions of power and benefits are the same, then S_2's payoff to aligning with the larger of A and S_1 is greater than its payoff to aligning with the smaller of the two states.

The Alignment Game

The alignment model is a two-stage game among the three players A, S_1, and S_2. The first stage determines which two of these states survives the first stage and move on to the second. The outcome of the first stage also determines what the distributions of power and benefits are between the two surviving states at the outset of the second stage.

Figure 5.1 describes the sequence of moves in the first stage. The states also begin the game with a distribution of territory q_A, q_1, q_2 where $q_A + q_1 + q_2 = 1$, and a distribution of power p_1^A and p_2^1 where $p_{\text{subscript}}^{\text{superscript}}$ is the probability that the state or coalition denoted in the superscript defeats and thereby eliminates the state or coalition in the subscript. (The probability that A defeats S_2, which is p_2^A, is not a primitive and is derived below from p_1^A and p_2^1.) Given these preliminaries, two things must be done to complete the specification of the first stage.

First, the probabilities of the random moves (i.e., the moves Nature makes) must be defined as a function of the initial distribution of power p_1^A and p_2^1. To derive these probabilities from p_1^A and p_2^1, it is convenient to think of each state as having some underlying military capability. Let k_A, k_1, and k_2 denote these capabilities where $k_A + k_1 + k_2 = 1$. Then the probability that state i defeats j is $p_j^i = k_i/(k_i + k_j)$, and the probability that state i prevails over the coalition composed of j and n is $p^i_{j,n} = k_i/[k_i + g(k_j + k_n)]$ where g measures the extent to which military capabilities cumulate.

To solve for p_2^A in terms of p_1^A and p_2^1, note that $p_j^i = k_i/(k_i + k_j)$ implies $k_i/k_j = p_j^i/(1 - p_j^i)$. Then eliminating k_A and k_2 from $p_2^A = k_A/(k_A + k_2)$ yields:

$$p_2^A = \frac{p_1^A p_2^1}{p_1^A p_2^1 + (1 - p_1^A)(1 - p_2^1)}.$$

It is also easy to find $p^i_{j,n}$ (and therefore $p_i^{j,n}$ as $p_i^{j,n} = 1 - p^i_{j,n}$). To wit,

$$\begin{aligned} p^i_{j,n} &= \left[\frac{k_i + g(k_j + k_n)}{k_i}\right]^{-1} \\ &= \left[1 + g\left(\frac{k_j}{k_i} + \frac{k_n}{k_i}\right)\right]^{-1} \\ &= \left[1 + g\left(\frac{p_i^j}{1 - p_i^j} + \frac{p_i^n}{1 - p_i^n}\right)\right]^{-1} \\ &= \left[1 + g\left(\frac{1 - p_j^i}{p_j^i} + \frac{1 - p_n^i}{p_n^i}\right)\right]^{-1} \\ &= \frac{p_j^i p_n^i}{p_j^i p_n^i + g(p_j^i p_n^i + (1 - p_j^i)(1 - p_n^i))}. \end{aligned}$$

The last task to be done in defining the first stage is to describe the possible outcomes of this stage. There are two kinds of outcomes. If the game ends in the first stage because A does not attack or because one state eliminates the other two states, then the states' payoffs have to be specified. If, by contrast, two states survive the first stage, then the distributions of power and benefits between these two states must be characterized. These distributions constitute the initial conditions under which the second-stage bargaining occurs. It will be convenient to represent this outcome by the four-tuple $(n, m, \widehat{q}_n, \widehat{p}_n)$ where n and m are the two surviving states, $\widehat{q}_n$ is state n's share of the territory at the outset of bargaining ($\widehat{q}_m = 1 - \widehat{q}_n$), and $\widehat{p}_n$ is the probability that n prevails if the bargaining breaks down in war.

If A does not attack, the game ends and the distribution of benefits is unchanged. The states' payoffs in this case are q_A, q_1, and q_2. Now suppose that A attacks both S_1 and S_2. If A prevails, it eliminates both S_1 and S_2 and the game ends. A obtains all of the territory and receives a payoff of one less the cost of fighting, i.e., $1 - c_A(q_1 + q_2)$. S_1, by contrast, retains control of nothing and pays the cost of fighting. Its payoff is $-p_2^1(c_1 q_A)$ where $c_1 q_A$ is the total cost S_1 would bear if it fought A alone. Similarly, S_2's payoff if A eliminates it is $-(1 - p_2^1)c_2 q_A$.

If, by contrast, A loses when it attacks both S_1 and S_2, the first stage ends and S_1 and S_2 move on to the second. By assumption, when two states fight together and eliminate a third, the distribution of power between the two victors after the war is presumed to be the same as it was before the war. Consequently, the distribution of power between S_1 and S_2 at the start of the second stage is p_2^1. We also assume that the fighting that eliminates one state divides that state's territory in proportion to the distribution of power between the two victorious states. Consequently, S_1's share of the territory at the end of the first stage is its original share plus what it captures from A: $q_1 + p_2^1 q_A$. S_2's share is $q_2 + (1 - p_2^1)q_A$. If, therefore, A attacks both S_1 and S_2 and loses, then the four-tuple $(S_1, S_2, q_1 + p_2^1 q_A, p_2^1)$ defines the conditions for the subsequent bargaining.

If A attacks only one of the other states, say S_i, then the outcome of the first stage depends on what S_j $(j \neq i)$ does. If S_j waits and A wins, then A obtains all of S_i's territory and acquires all of its military capabilities. The resulting outcome is the four-tuple $(A, S_j, q_A + q_i, p_j^{A,i})$. If S_j waits and S_i prevails, the outcome is $(S_i, S_j, q_A + q_i, p_j^{A,i})$.

Suppose instead that S_j aligns with A when A strikes S_i. Then the game ends if S_i prevails, and S_i captures all of the territory at cost $c_i(q_A + q_j)$. If the coalition of A and S_j prevails, then A and S_j divide S_i's territory and the distribution of power between them remains p_j^A. The outcome is $(A, S_j, q_A + p_j^A q_i, p_j^A)$.

Finally, assume A attacks S_i and S_j aligns with S_i. If A prevails, the game is over. If S_i and S_j triumph, then the second stage begins with $(S_i, S_j, q_i + p_j^i q_A, p_j^i)$.

If two states survive the first stage, then their payoffs depend on the outcome of the bargaining, which depends on the initial distributions of power and benefits between the bargainers. Suppose that the alignment stage results in the four-tuple $(n, m, \widehat{q}_n, \widehat{p}_n)$. Then in keeping with the bargaining model discussed in chapter 3, n is dissatisfied if its payoff to fighting is greater than its payoff to living with the status quo: $\widehat{p}_n - c_n\widehat{q}_m > \widehat{q}_n$. Similarly, m is dissatisfied if $\widehat{p}_m - c_m\widehat{q}_n > \widehat{q}_m$. (As before, both states cannot be dissatisfied at the same time.)

The payoffs associated with $(n, m, \widehat{q}_n, \widehat{p}_n)$ depend on whether or not one of the states is dissatisfied. If n is dissatisfied, then we assume that m concedes just enough territory to n to make it indifferent between attacking and accepting the offer. In symbols, n offers x where $x =$

$\widehat{p}_n - c_n\widehat{q}_m$. The payoffs associated with $(n, m, \widehat{q}_n, \widehat{p}_n)$ if n is dissatisfied therefore are $\widehat{p}_n - c_n\widehat{q}_m$ for n and $1-(\widehat{p}_n - c_n\widehat{q}_m)$ for m. If, by contrast, m is dissatisfied, then m's payoff is $\widehat{p}_m - c_m\widehat{q}_n$ and n's payoff is $1 - (\widehat{p}_m - c_m\widehat{q}_n)$. If both states are satisfied, neither can credibly threaten to use force to revise the status quo. In this case, n's payoff is $\widehat{q}_n$ and m's is $1 - \widehat{q}_n$.

The definitions of the payoffs associated with $(n, m, \widehat{q}_n, \widehat{p}_n)$ complete the specification of the alignment game. This perfect-information game is readily solved by backward programming, and the results of doing so for different parameter values are reported in chapter 5. However, two related remarks about different ways of thinking about the payoffs associated with $(n, m, \widehat{q}_n, \widehat{p}_n)$ are in order.

First and most directly, we can simply take these payoffs to be part of the definition of the game. In this interpretation, there is no formally specified bargaining game in the second stage. Rather, the bargaining game discussed in chapter 3 is only used informally to motivate the definition of the states' payoffs.

Alternatively, we may conceive of the payoffs associated with $(n, m, \widehat{q}_n, \widehat{p}_n)$ as those that actually do result from solving a well-defined bargaining game. Indeed, we might use the bargaining game discussed in chapter 3 and formally analyzed in appendix 3 in this way. Although this can be done, there is an important difference between the assumptions underlying that game and those underlying the alignment game. The former assumed that the cost of fighting was constant, whereas the cost of fighting in the alignment game depends on an adversary's size. Because of this difference, the bargaining game in chapter 3 provides an unnatural model of the bargaining in the alignment game.

In chapter 3, the bargaining and the game end as soon as one of the states accepts an offer. There are no further opportunities to revise the distribution of territory. This assumption simplifies the model and the characterization of the equilibria. But, we can easily imagine a more complicated bargaining game in which offers and counter-offers can continue as long as no state has attacked. In this formulation, the distribution of territory could be revised multiple times.

As it happens, the solutions to the simple game and to the more complicated one are the same if the cost of fighting remains constant.[1] Recall what happens in the simple game in which the bargaining ends as soon as

[1]This equivalence also presumes that the satisfied state makes the first offer. Matters are somewhat more complicated but essentially the same if the dissatisfied state makes the first offer and the discount factor is close to one.

a proposal is accepted. With complete information, the satisfied state offers the dissatisfied state its certainty equivalent for fighting, and the dissatisfied state accepts. This revision in the distribution of territory leaves both states satisfied. Accordingly, the distribution of territory would not be amended any further even if the bargaining could continue.

If there were asymmetric information in the simple game, then the satisfied state would make its optimal take-it-or-leave-it offer. The dissatisfied state then either attacks or accepts depending on whether or not its payoff to fighting is greater than its payoff to accepting the offer. Because the only types that accept an offer are those that prefer it to fighting, neither state would be willing to use force to change the distribution of territory again if the first proposal is accepted. Thus there would be no subsequent revisions even if the states could continue to negotiate. It does not matter whether the bargaining can continue or not if the cost of fighting (or more generally the payoff to fighting) remains constant.

But it does matter whether or not the bargaining can continue if the cost of fighting depends on an adversary's size. If the bargaining ends as soon as a proposal has been accepted, then the status quo will be revised only once. If, by contrast, the bargaining can continue as long as neither state attacks, then the status quo will be revised infinitely often.

To see that this is so, suppose the bargaining ends as soon as one state accepts an offer. Then the satisfied state in this formulation again offers the dissatisfied state its certainty equivalent. In symbols, the satisfied state grants the dissatisfied state $x^* = p - d(1 - q)$ where p is the probability that the dissatisfied state prevails, d is its marginal cost of fighting, and $1 - q$ is the status quo size of the satisfied state.

Now assume that the bargaining can continue until one of the states attacks. The satisfied state no longer makes one large concession equal to x^*. Rather, it makes a series of concessions $x_0, x_1, x_2, \ldots$ where the first offer x_0 is smaller than x^* and the series of concessions converges in the limit to $(p-d)/(1-d)$. It is easy to verify that this limit is larger than x^* as long as the dissatisfied state is actually dissatisfied (i.e., as long as $p - d(1 - q) > q$). Thus the initial offer x_0 is less than what would have been offered if the bargaining stopped after one revision (i.e., $x_0 < x^*$) and the latter offers are larger than what would have been offered if the bargaining stopped after one revision (i.e., $x^* < (p - d)/(1 - d)$).

To derive these concessions, consider the satisfied state's first offer x_0. If the dissatisfied state rejects this offer and fights, it obtains $[p - d(1 - q)]/(1 - \delta)$. If, however, the dissatisfied state accepts x_0, it can always fight in the next period when its cost will be less because its adversary will be smaller. The payoff to doing so is $x_0 + \delta[p - d(1 - x_0)]/(1 - \delta)$. The satisfied state can therefore induce the dissatisfied state to postpone an attack for at least the current period by offering an x_0 such that

$$\frac{p - d(1-q)}{1-\delta} = x_0 + \frac{\delta(p - d(1 - x_0))}{1-\delta}$$
$$x_0 = \frac{(1-\delta)(p-d) + dq}{1 - \delta + \delta d}$$

More generally, the satisfied state can get the dissatisfied state not to attack at time t when the distribution of territory is x_{t-1} by conceding $x_t = [(1-\delta)(p-d) + dx_{t-1}]/(1 - \delta + \delta d)$. A straightforward application of the proof of proposition 1 in Fearon (1996) shows that x_t is in fact the equilibrium offer of an underlying bargaining game in which the satisfied state makes all of the offers and the bargaining continues as long as neither state attacks.[2]

Thus, whether or not the bargaining can continue after a proposal has been accepted affects the pattern of concessions when the payoff to fighting is a function of past concessions. But, interestingly, it does not affect the states' payoffs in the game. The dissatisfied state's payoff to receiving $x_0, x_1, x_2, \ldots$ is just $\sum_{t=0}^{\infty} \delta^t x_t$. It is easy, if tedious, to solve for x_t recursively for q to show:

$$x_t = \sum_{j=0}^{t}\left[\left(\frac{d}{1-\delta+\delta d}\right)^j \frac{(1-\delta)(p-d)}{1-\delta+\delta d}\right] + \left(\frac{d}{1-\delta+\delta d}\right)^{t+1} q$$
$$= \frac{p-d}{1-d}\left[1 - \left(\frac{d}{1-\delta+\delta d}\right)^{t+1}\right] + \left(\frac{d}{1-\delta+\delta d}\right)^{t+1} q$$

Substituting this expression for x_t in $\sum_{t=0}^{\infty} \delta^t x_t$ gives

$$\sum_{t=0}^{\infty} \delta^t x_t = \frac{p - d(1-q)}{1-\delta}.$$

But this payoff is just what the dissatisfied state would have received if the bargaining had ended as soon as a state accepted an offer. As we have seen, the satisfied state would offer $x^* = p - d(1-q)$ in those circumstances, and the dissatisfied state's payoff would be $[p - d(1 - q)]/(1-\delta)$. (An intuition for the equivalence of these payoffs is that regardless of whether or not the bargaining continues, the satisfied state obtains all of the surplus saved by not fighting. Whether or not the bargaining continues only affects the way that this surplus is collected.)

In sum, we can think of the payoffs associated above with the outcome $(n, m, \widehat{q}_n, \widehat{p}_n)$ as the average payoffs of the states in the rather unnatural bargaining game in which the interaction stops as soon as an offer is accepted. Or, we can take the payoffs associated with $(n, m, \widehat{q}_n, \widehat{p}_n)$ to be the average payoffs of a bargaining game in which the bargaining can continue as long as no state attacks.

[2] Fearon studies a model in which the distribution of power is a function of the distribution of territory. Thus, a concession to an adversary in the current period makes that adversary stronger in the next, and this increases the adversary's payoff to attacking. The analysis of that model carries over to the present situation in which a concession in the current period reduces an adversary's cost of fighting in the next and thereby raises its payoff to attacking.

Aligning with Larger States

Consider now the claim that if the distribution of power mirrors the distribution of benefits and the potential attacker actually attacks one of the other states, then the third state's payoff to aligning with the larger and therefore more powerful of the other two states is greater than its payoff to aligning with smaller of the other two states. To establish this, suppose A attacks S_1. Then the claim asserts that S_2's payoff to aligning with the larger of A and S_1 is greater than the payoff to aligning with the smaller of those two states.

Because the distribution of power mirrors the distribution of benefits, S_2 and its coalition partner will be satisfied with each other if they move on to the bargaining stage. This means that S_2's payoff to aligning with A is $p_1^{A,2}(q_2 + p_A^2 q_1) - c_2 p_A^2 q_1$. Similarly, S_2's payoff to aligning with S_1 is $p_A^{1,2}(q_2 + p_1^2 q_A) - c_2 p_1^2 q_A$.

The fact that the distributions of power and benefits mirror each other also means that we can treat each state's territory as a measure of its military capability, i.e., $k_A = q_A$, $k_1 = q_1$, $k_2 = q_2$. Substituting these into the expression for S_2's payoff to aligning with A gives:

$$\frac{g(q_A + q_2)}{q_1 + g(q_A + q_2)}\left(q_2 + \frac{q_2 q_1}{q_A + q_2}\right) - \frac{c_2 q_2 q_1}{q_A + q_2}$$

Substituting the expressions for the states' capabilities into the expression for S_2's payoff to aligning with S_1 and recalling that $q_A + q_1 + q_2 = 1$ show that the difference between S_2's payoff to aligning with S_1 and A is:

$$(q_1 - q_A)\left[\frac{g(1-g)}{[q_A + g(1-q_A)][q_1 + g(1-q_1)]} + \frac{c_2}{(1-q_A)(1-q_1)}\right]$$

Thus, S_2's payoff to aligning with S_1 is larger than its payoff to aligning with A if and only if $q_1 > q_A$.