since the measuring device has been constructed by the observer . . . we have to remember that what we observe is not nature itself but nature exposed to our method of reasoning.
—werner heisenberg, 1958

In the previous chapter, we laid out the epistemological issues surrounding our research design and many of the choices we made therein. Next, we discuss in detail the models, arguments, and conjectures that we will examine with our statistical tools. We group these by level of analysis, beginning with the individual state level. We then turn to the dyadic level, or pairs of states, and finally examine the arguments cast at the international system level. In chapter 6 we look at regional level variation. For each argument we test, we first lay out the logic or intuition behind it. Then we explain how one might go about testing empirical propositions consistent with the argument’s theoretical expectations. To accomplish this, we operationalize the set of key concepts that each hypothesis implies should be associated with more or less international conflict. In many instances, the logic behind the “theory” is so vague or poorly specified that it is impossible to test critical underlying causal steps of the argument. Instead, we set goals that are more modest—we aim to test carefully the empirical claims various authors make about the relationship between operational measures linked to their stories and the onset and escalation of interstate conflict. Part of the dilemma is that, in most cases, the various authors do not really have an explicit theory of the decisions that lead to war. Instead, they make conjectural arguments linking some operational measure to some vaguely specified mechanism that makes the decisions leading to war more or less likely. Our aim here is to see which of these conjectures finds enough
empirical support to suggest where further theoretical development will be most fruitful.

STATE LEVEL OF ANALYSIS

In this section, we discuss our indicators that purport to measure the effects of state level, or monadic, factors on the relative likelihood of war. The arguments in this section predict the behavior of states in a dyad based solely on a single state’s internal characteristics, without reference to the joint characteristics or interactions within the dyad.

1. Democratization

While most scholars agree that democracies fight less with one another than with other types of states, Mansfield and Snyder (1995) argue that the transition to democracy is quite dangerous, largely because of institutionally insecure leaders looking for support from an increasingly nationalistic populace. Others disagree with this proposition (Thompson and Tucker 1997; Enterline 1996, 1998; Ward and Gleditsch 1998). Some critics concentrate on Mansfield and Snyder’s research design, changes in which seem to yield rather different conclusions about the dangers (or lack thereof) of democratization. Others focus on the relatively simplistic initial conceptualization and measurement of what constitutes democratization. We focus on Ward and Gleditsch’s work (1998), which presents one of the more sophisticated analyses of whether and how democratization affects international conflict. Ward and Gleditsch break the concept of democratization into three separate aspects of regime change: direction, intensity, and nature. By “direction,” they mean whether the change in a regime is toward or away from democracy. By “intensity,” they mean how rapidly change occurs or how much of a shift in the nature of the regime there ultimately is. Last, by the “nature” of regime change, they address whether the shift is smooth and relatively linear or if it is marked by reversals or rapid spurts in regime transformation. It is this third characteristic in particular that is unique to their analysis.

We follow the operational methods detailed in table 2 of Ward and Gleditsch’s (1998) work but use updated data and make minor modifications following conversations with Ward and Gleditsch. Using the Polity IV data, we first compute the change in a state’s regime type over a ten-year period as \((\text{Democracy}_{i,t} - \text{Democracy}_{i,t-10})\). If the tenth lag
is missing, we use the next closest possible lag, a nine-year lag if available, otherwise an eight-year lag, and so forth. We also include the direction of that regime change as a second variable, which takes the value $-1$ if change was away from democratic regime characteristics, $0$ if there was no change, and $+1$ if change was toward democratic regime characteristics. Finally, we include a measure of the variance of change in the state over the ten years. If data were missing during the period, we used the variance of the available data, as long as there were at least three years of data. We include these three variables separately for each state in the directed dyad. We are then able to observe whether democratization in the potential initiator, potential target, or both has an associated effect on conflict behavior.

2. Polity Change and Externalization of Violence

Related to the democratization literature is a set of arguments about regime change in general. Levy (1988) summarizes these arguments and points out that a government’s concerns about maintaining internal support may lead to war through any of several mechanisms. Analysts of externalization and diversionary war argue that conflict is more likely when states face significant internal problems. Problems such as economic downturns and a subsequent (or anticipated) drop in internal support may lead national leaders to look for ways to boost their domestic political standing by searching for an explicit “rally round the flag” effect (Russett 1990). Historically, appeals to nationalism have served states well and often serve as sufficient justification to generate internal support for war (Snyder 2000; Reiter and Stam 2002). Following the rally round the flag logic, military confrontation with another state may provide a way to achieve a boost in popular support for a state’s leader or support for a larger war that the public would not have otherwise backed. Similarly, when there is a change of government in a state, a new government may look for ways to bolster internal support, as well as to establish an international reputation for toughness, which state leaders sometimes focus on to the detriment of long-term state goals (Mercer 1996). In this latter situation, other states may see an opportunity to press a claim over some disputed issue because they see the government as less in control than an established regime (Blainey 1988). According to the logic of either argument, a state’s involvement in international conflict will be more likely when there is substantial change in a state’s domestic political institutions.

An alternative chain of logic in the psychological literature focuses
on leaders’ reactions to crises. These arguments suggest that, in situations of internal political vulnerability, leaders may suffer from motivated bias, leading them to see the enemy as they wish to see them, typically weaker and making an easier target than would be expected given a “rational” evaluation of the situation (Jervis 1976; Lebow 1981). Expectations about the escalation of conflict under these conditions are less than clear, however. Following the diversionary argument from a rationalist unitary actor perspective, states engaged in actions to shore up their domestic political situation may not want to pay the costs of a long-running conflict or war, and so we should see less escalation beyond low-level disputes. However, if challenged by another state, leaders facing intense internal political dissent may be likely to force a confrontation initially in order to gain needed, and expected, domestic political benefits by standing tough against an external foe. Similarly, if such a state faces a challenge over some outstanding issue, there is no guarantee that it will be more willing to give in to the challenger compared to times when the government is in a position of domestic political strength. Following the motivated bias argument, we might expect more or faster conflict escalation as a leader blindly pushes ahead in his or her dispute with an opposing state.

Insufficient systematically collected historical data make it impossible for us to test directly arguments about the effects of internal political conditions using measures of a state’s economic situation or the relative number of political protests in a cross-national time-series analysis. However, we can examine the effects of abrupt changes in a regime’s political institutional characteristics on international conflict. Based on Polity IV data we include a variable that marks whether the states in a dyad experienced an abrupt polity change. As noted in the democratization section, we use several variables to code general polity changes. We also mark a change corresponding to two additional conditions: (1) when a regime comes to an abrupt end according to the Polity IV “eyear” variable and (2) under conditions marking a polity change coded when the “polity” variable (autocracy-democracy) takes a value of \(-66\), \(-77\), or \(-88\), indicating an interruption, interregnum, or transition period, respectively. All of these are circumstances where the state’s political institutions are in significant flux. We lag the regime change variable one-year to avoid the possibility of accidentally picking up conflicts that led to a polity change rather than vice versa. If arguments about domestic instability leading to international conflict initiation or escalation are correct, following a polity change we should find that (1) states initiate disputes more frequently but may escalate those disputes less often and/or (2) states initiate and/or escalate disputes
Next, we turn to arguments cast at the dyadic level of analysis, beginning with alliances and defense pacts.

3. Alliance and Defense Pact Membership

Alliances lie at the core of a vast literature. These foci include the realist balance of power/threat debate (Morgenthau 1956, Walt 1987), empirical work in international relations (Singer and Small 1967), the formal or deductive rational choice literature (Bueno de Mesquita 1981), and the norms or domestic political institution literature (Risse-Kappen 1991; Simon and Gartzke 1996). The best available evidence suggests that military alliances between states serve several functions. Early empirical work demonstrated their role as mechanisms of war diffusion (Siverson and King 1980; Siverson and Tennefoss 1984; Oren 1990) and as capability-aggregation mechanisms (Morgenthau 1956; Liberman 1996; see Brooks 1998 for an alternative view). More recently, others argue that they serve as indicators of bilateral satisfaction (Lemke and Reed 1996) and as a sign of beliefs about the likelihood and nature of future conflict (Walt 1984; Morrow 1999). More generally, alliances may be simple reflections of common interests (Bueno de Mesquita 1981). Longer-term interests and continuing interaction may lead allies to trade more with one another, or they may develop certain shared institutional structures over time that will provide incentives and/or mechanisms to avoid conflict. We do not attempt to sort out the often conflicting logic of the many hypotheses about the effects of alliances on international politics. These include how alliances serve states’ interests and how they affect the nature of interstate conflict (for a large set of additional inductive hypotheses see Holsti, Hopmann, and Sullivan 1972). Our tests here simply aim to identify whether and when alliances are associated with the onset and escalation of interstate conflict.

We include two measures of dyadic alliance membership to see whether allied states are better able to avoid conflict than nonallied states. Alliances may take many forms: offensive, defensive, ententes, and formal institutionalized arrangements (such as NATO). We first
include a variable marking whether a formal defense pact existed between the members of the dyad in the given year. To capture the basic intuitions of those that argue that alliances simply aggregate capabilities and those that argue that some alliances may evolve into theoretically distinct defense communities, we include a separate measure marking that the two states were jointly members of NATO. This second measure is particularly important to control for a confounding alternative to the democratic peace argument, namely, the possibility that in fact it is NATO rather than democracy that has driven peaceful relations in Europe since World War II and provided evidence for the apparent democracy-peace relationship. The argument points out that the most stable democracies since World War II have also been members of NATO. If NATO is the true motive force behind peace between these countries, then the democratic institutions’ contribution is merely an illusion.

4. Arms Races

Arms races occur when two states build their military armaments in response to each other’s purchases and manpower mobilizations, leading to an action-reaction cycle of rapidly increasing military expenditures and/or forces. Some of the most sophisticated and careful thinking about the nature of arms races harks back to Huntington’s (1958) and Richardson’s (1951, 1960a, 1960b) work in the 1950s. Huntington observed that an arms race could take place both in terms of the quantity of arms and in terms of technological innovations (see also Evangelista 1988). Richardson developed the first mathematical formalizations of arms races in the 1930s, models that remain the foundation for much current formal work in the area.

The argument that arms races increase the probability of conflict and escalation draws more recently from Jervis’s (1976) informal but highly influential spiral model of hostility (see also Herz 1959; Butterfield 1951). According to this model, as each state observes the other building up its weaponry and reacts in-kind, the level of hostility between the states increases, because neither side can know the true intentions of its ever increasingly armed potential opponent. The key factor lies in each side’s uncertainty over the private intentions of their opponent (Kydd 1997). It is difficult for a state to claim credibly that their increased arms are strictly for defensive purposes. According to Jervis, arms races create an intensifying cycle of hostile action and reaction, which ultimately may lead either to inadvertent or to deliberate armed conflict. The increase in
hostility and the expected subsequent increase in conflict are consistent with both rational and cognitive psychology models. In the rational explanation, the observation of an opponent’s buildup is a signal of hostile intent on the part of the other actor. In psychologically based models, arms races produce misperception, which leads to tension, and increase the likelihood of emotionally driven or expressive reactions (as opposed to instrumentally rational choice), which in turn may lead to unintended conflict escalation.

An important debate in the empirical literature broke out in the early 1980s over whether increases in arms did in fact increase the probability of conflict. Wallace (1982) found that arms races had a major substantive effect associated with an increase in likelihood of war, while Diehl (1983) found that arms races had little, if any, impact on the probability of war (see also Intriligator and Brito 1984; Diehl 1985). Subsequent formal analysis by Morrow (1989) suggests that a more sophisticated arms race model based on a rational choice premise can help to explain the difference in findings. The most recent empirical work on this subject suggests that arms races do in fact contribute positively to dispute escalation (Sample 1997). However, Sample’s work fails to control for the potentially confounding effects associated with power transition logic, which in many instances should be quite collinear with arms races. While the measures may be quite collinear, the two arguments (arms races and power transitions) rest on quite different theoretical foundations. Arms races focus on military expenditures and joint reactions therein. Power transition logic focuses on the relationship of more aggregate measures of state power, status hierarchies, and relative status or dissatisfaction. One could argue, however, that in both cases the measures marking the presence of an arms race are poor instruments for the underlying theoretical mechanisms, which in fact are tapping into the same dyadic pressures. The aggregate data approach we employ here cannot distinguish between the psychological and signaling variants about how arms races affect interstate conflict. However, we do control for other factors, such as the presence of systemic and dyadic power transitions, and so hope to establish more clearly the empirical relationships between carefully specified operational measures of arms races and conflict.

An empirical criticism of using any objective measure of arms races notes that there may be multiple reasons or motives for what appears to be “arms racing,” such as bureaucratic incentives to expand budgets or technological pressures to innovate (see Evangelista 1988 for a discussion of the role of technological innovation compared to quantitative arms racing). However, subjective measures of when leaders perceived themselves to be in arms races do not exist.
For our logit model, we build on Diehl’s (1983) measure of arms races and code an arms race as present if the three-year moving average of constant-dollar military expenditure growth is greater than 8 percent for both states. Diehl’s measure used current military spending; here we deflate spending to avoid coding as present an arms race simply because one or both states are experiencing a period of high inflation. One could argue from a psychological perspective (Jervis 1976) or a dominant indicator approach (Gartner 1997) that leaders attend to other measures, potentially including actual (current) spending along with stocks of particular weapons or specific technological innovations. As one check, an analysis using Diehl’s measure based on current expenditures produces similar results to those we report here. Nevertheless, in keeping with existing arguments and tests, we maintain our focus on military expenditures.

We include only the three years before the occurrence of a militarized dispute and not expenditures for the year of the dispute. If the moving averages show greater than an 8 percent average military expenditure growth for both states, we code a final dummy variable marking arms race as a “1.” If this joint condition does not hold, we code a “0.”

5. Balance of Power

The notion that an equitable balance in power or capabilities will maintain the peace between nations is one of the oldest and most central notions in world politics. It is also one of the most hotly contested. At the most basic level, balance of power theorists maintain that when a balance of military power, typically a relatively equal one, exists between two or more states, the states in question will be less likely to go to war with one another (Morgenthau 1956; Waltz 1979). Waltz argues explicitly that it is the likelihood of high costs and uncertain outcomes that exist when power is equitably balanced that leads to stability and peace, suggesting that “Where a balance of power does exist, it behooves the state that desires peace as well as safety to become neither too strong nor too weak” (Waltz 1959, 222). In contrast, when one state is more powerful than another, it may go to war against the other to enhance further its power position; in this view, weakness thus encourages aggression.

The balance of power literature is too extensive to allow a detailed review in the space of this entire volume, much less a section of a chapter (see Levy 2002). Realism, the international relations paradigm based
on states’ pursuit of power or security in an anarchic environment, has the most to say about the role that balances of power should play in interstate relations. Perhaps the earliest recorded balance of power theorist was Thucydides. Laying the foundations for much of the modern study of international politics, Morgenthau (1956) and Kaplan (1957) laid out what were, in essence, inductive arguments linking relative power parity to the likelihood of conflict. One source of confusion regarding balance of power arguments has to do with what precisely authors mean when they refer to the, or a, balance of power. Some refer to balancing behavior in the context of a state’s option to join potential alliances (the alternative being bandwagoning), while others such as Waltz argue that the balance of power explanation of war “is a theory about the results produced by the uncoordinated actions of states” (1979, 122). Waltz’s characterization is one of the clearest linking the behavior of states to some balance of their capabilities. In an elegant formal model of how power balances might affect the probability of war, Powell (1999) argues that “War is least likely when the existing distribution of benefits reflects the underlying distribution of power” (85). As he points out, this argument is in contrast to the standard supposition of the balance of power school, that power parity between actors should be associated with the lowest probability of war.

Unfortunately, testing balance of power “theory” in its common variants poses several problems. These problems interfere with developing a test of the argument as a whole for two principal reasons. First, several variants have been developed that define the concepts of “balance” and “power” in different ways. For instance, power may flow from capabilities (which are measurable ex ante) or on outcomes (in which case the argument becomes tautological). Some have noted that the “balance of power” may be a description of the status quo, a description of past conflict behavior, or a description of a state’s foreign policy. Even once a conflict begins, the “balance of power” may exist at different levels. For example, in a simple model of bargaining and war, Smith and Stam (2001) show that the “balance of power” exists meaningfully in at least three separate ways. They point out the following interpretations of the “balance of power:” (1) the likelihood of one side or the other winning a single battle; (2) the likelihood of one side or the other being able to win a war to the finish; or (3) either side’s subjective beliefs about either of the first two.

A second problem is that some variants of “balance of power” logic are inconsistent (see Organski 1958; Zinnes 1967; Bueno de Mesquita 1980; Niou, Ordeshook, and Rose 1989). For example, in contrast to the standard realist notion that an equal balance of power between two
states (commonly assumed to mean the bilateral or dyadic military balance of capabilities) will reduce the likelihood of conflict, an equally logical argument has been made that states should in fact be most likely to fight when power is equal. According to the “power preponderance” model of war, when capabilities are roughly equal, the states involved will be less certain of a potential war’s outcome than when there is a clear advantage. They may then both “push” in a dispute or low-level conflict, each believing it can prevail. As a result, low-level conflicts will be more likely to escalate into larger ones (Blainey 1988). However, when one side enjoys a preponderance of power, the outcomes of potential conflicts are clear and states will settle disputes before they escalate to war. According to this logic, it is when both states see a reasonable chance of winning—namely, when power is balanced—that disputes will be most likely to break out and to escalate to war.

We do not test various ancillary arguments and extensions to the basic balancing logic. Therefore, we set aside questions about how states react to shifting power balances depending on their risk attitudes or their relative degree of satisfaction (Huth, Bennett, and Gelpi 1992; Lemke 2002). Many critiques of balance of power approaches suggest that it is too simplistic to expect all situations of balance to provide the same incentives for states to react uniformly and suggest interacting power balances with markers for particular situations when an equal power balance should lead to peace. Much of the recent work in this area has focused on how uncertainty about the bilateral balance of power may drive conflict, focusing on incomplete or private information about costs, outcomes, and durations. Most authors cite Blainey (1988) as the progenitor of this literature, primarily focused on bargaining and war (e.g., Fearon 1996; Wagner 2001; Werner 2000; Powell 2002; Reiter 2003; Smith and Stam 2001). While this is one of the most exciting areas of theory development concerning the role of national power to emerge in the past several years, we do not provide direct tests of related hypotheses here. It is not clear yet how to test these theories directly since we, as analysts, are likely to share the same ex ante information asymmetries or deficits as do the decision makers.

While direct tests of models driven by private information are beyond the scope of this project, we do provide some indirect tests of the balance of power approach. Some of the operational indicators claimed by various scholars to be direct measures of the critical variable needed to “test” their “theory” may actually be serving as proxies or indirect measures of incomplete information or uncertainty (for example, system power concentration or arms races may influence, or reflect, the level of uncertainty in the system). It may be this uncertainty that drives the
observed probability of mutual armed conflict. Ultimately, the constraint of being unable to measure uncertainty directly may not be a problem if we can develop theoretically sound instruments or proxies. However, the challenge of integrating various interpretations of such measures leads us to the conclusion that we need more and better theory and more thought behind what our various measures actually capture rather than simply more rigorous empirical testing of existing arguments.

In our analysis that follows we ignore the logical flaws in balance of power arguments and instead focus on an empirical proposition drawn from one of the most basic balance of power notions, namely, that the closer the capabilities of two states are to one another, the lower the probability is of conflict initiation or escalation. We measure “power” using the national capability score from the COW project (Singer, Bremer, and Stuckey 1972). We create a ratio of the larger state’s capabilities to the total capabilities of the dyad. The final variable ranges from $0.5$ (at parity) to $1.0$ (when one state possesses 100 percent of the capabilities in the dyad). Balance of power advocates would expect the variable to relate positively to conflict, while power preponderance advocates predict the opposite. Note that while the underlying argument whose hypothesis is being tested here is balance of power “theory,” the variable in fact measures the balance of capabilities; we refer to the variable as the “balance of forces” to maintain the operational distinction.

6. Conventional Deterrence

Conventional deterrence models focus on the potential costs, either human or material, faced by or threatened against a potential attacker that may dissuade it from using force against a deterring state or its protégé, logic somewhat similar to balance of power arguments. In a setting of immediate deterrence, a variety of factors may come into play (see Huth 1988; Huth, Gelpi, and Bennett 1993; Huth and Russett 1993; Mearsheimer 1983). Here, in our general deterrence setting, we examine the associations between military superiority and the potential attacker’s probability of initiating a conflict as compared to the non-directed hypotheses that emerge from balance of power logic. When considering conventional warfare, some scholars argue that deterrence by denial is likely to be an effective means of deterring a potential attacker. If an attacker’s chances of victory decrease, or if the costs of likely victory increase, then a rational attacker is more likely to demure.

Frequently, scholars model the attacker’s naive probability of win-
ning simply as the attacker’s capabilities divided by the dyad’s total capabilities. Such a measure necessarily neglects the possible contributions of allied forces and distances between the attacker and the defender. To incorporate at least some of these contributing factors to likely outcomes of a conflict, we include in our model the initiator’s subjective probability of winning a conflict developed as a component of the expected utility measures underlying the IIG predictions. This subjective probability discounts the direct capabilities of the initiator and target in a dyad by distance and adds expected assistance for each side by examining the likely contribution of other states in the international system for each of the dyad members. Using the measure in this way does not run the risk of introducing significant collinearity with the IIG equilibria variables, because the probability measure filters through several additional steps in leading to the dummy IIG predictions. Here, we include our measure of predicted war outcomes (predicted probability of winning from the initiator’s point of view) as a continuous measure ranging between zero and one.

It is important to note that the measure is expected to be correlated with the balance of power measure discussed previously. At its heart, our war outcome probability measure remains the (adjusted and supplemented) capabilities of a potential attacker divided by the (adjusted and supplemented) summed capabilities of both states in the dyad. The balance of power measure is the (raw) capabilities of the larger state divided by the (raw) summed capabilities of both states. When the potential initiator is the larger state, these measures should have similar values. The two measures are theoretically distinct, however, and, empirically, differences created by adjusting capabilities in the probability of winning measure and in those dyads where the target is larger lead to sufficient differential variance to compute reliable estimates of their separate effects.

7. Democratic Peace

The large body of literature on the so-called democratic peace derives from the straightforward empirical observation that there has not been a war between modern liberal democracies. Importantly, the democratic peace proposition (DPP) has taken root in U.S. foreign policy circles; President Bill Clinton cited the proposition as partial justification for the U.S. military involvements in Haiti and the Balkans. George Bush noted the potential democratization of Iraq as a motive for a U.S.-led war in 2003. Recent theoretical arguments about the relationship
between democracy and war begin with Doyle (1983), who built on Kant’s notion of an emergent perpetual peace among liberal states based on shared norms of democratic compromise and the presence of restraining domestic institutions that prevent democracies from fighting one another. Following shortly after Doyle, Russett (1990) and Maoz and Russett (1993) focused on the role of elite level norms, or standards of behavior. More recently, rational choice scholars turned their theoretical lenses on variants of an institutional constraint mechanism (Bueno de Mesquita et al. 2001; Reiter and Stam 2002) and signaling games (Schultz 2000). In the former, democratic leaders choose policies more carefully due to potentially punishing electoral constraints. In the signaling literature, democracies are better able to signal their resolve than other types of states. Empirically, Oneal and Russett (2001b) present some of the most thorough tests of the DPP, although the first observations of the relationship between joint democracy and war appeared earlier (Babst 1964; Small and Singer 1976).

The democratic peace literature contains two dominant strands. The first uses increasingly sophisticated statistical methodologies and controls to increase our certainty regarding the veracity of the empirical claim (e.g., Senese 1997; Beck, Katz, and Tucker 1998; Raknerud and Hegre 1997; Green, Kim, and Yoon 2001). The second begins with the premise that the empirical finding exists and seeks to develop better theoretical explanations of the supposed facts (e.g., Schultz 2000; Fearon 1994; Bueno de Mesquita et al. 1999; Cederman 2001). While most current research continues to suggest that jointly democratic dyads are much less conflict prone than others, arguments persist over alternative explanations of the DPP. The most common counter-arguments are that the “democratic peace” is merely a coincidence of alliances, the cold war, or some other factor creating common interests between democracies (e.g., Gowa 1999; Gartzke 1998; Lemke and Reed 1996; Henderson 2002). Other systematic empirical work has found that democracy does not exert the same pacifying effect for lower-scale dispute initiation as for wars (Senese 1997; Reiter and Stam 2003), suggesting a more complex relationship between democracy and conflict than a straightforward extension of the democracy and war findings might suggest. Finally, there is some evidence of an “autocratic peace,” and it appears that pairs of similar states (either autocratic or democratic) appear to fight less than mixed pairs (Werner 2000; Raknerud and Hegre 1997). This suggests that political similarity or shared identity (rather than shared democracy) is the key to conflict avoidance.

Unfortunately, most of these studies have not taken a comprehensive
approach when controlling for alternative explanations, including those suggested by expected utility models, which explicitly incorporate shared interests. Here we do not examine any particular mechanism through which democracy affects the level of interstate conflict. Instead, we examine whether the democratic peace finding holds up, given the most comprehensive controls to date. Ours tests include variables to examine numerous competing explanations and forms of the relationships between domestic political institutions and the likelihood of disputes and war.

Our specification for testing the effect of joint polity type on conflict is substantially more complex than previous designs, but it allows us to obtain more nuanced findings than existing work. We begin with the now-standard individual measures of regime type for the potential initiator and potential target in each dyad; we then add several interactions and variables measuring political similarity in order to explore a variety of hypotheses about democracy, democratization, and interstate conflict. As a result, we are able to capture the monadic effects of polity type on conflict, a variety of dyadic effects, the possibility of an autocratic peace, and the possibility that it is political similarity rather than shared democracy that drives the democratic peace.

We begin with individual regime-type scores from the Polity IV data set (see Jaggers and Gurr 1995; Marshall and Jaggers 2000). The Polity IV data now include this base regime-type coding as the “polity” variable, while other work has previously called it the “dem” score. The variable consists of the level of democracy in a state (“democ”) minus the level of autocracy (“autoc”). This index ranges from −10, indicating states with low democracy and high autocracy, to +10, indicating states with the opposite. Following the methods suggested in the Polity IV codebook we interpolate data for years in which there is an interruption or interregnum (polity codes −66, −77, −88), setting those years to the average of the last year before interruption and first year after interruption, as long as that gap was not greater than five years in either direction. This interpolation allows us to measure the “polity” score for almost five hundred more nation-years than we would be able to otherwise (of about twelve thousand) and political similarity in about twenty thousand additional dyad-years. In most cases, the interpolation fills in a gap of one to three years of data; in no case does it fill in more than five years sequentially for a single country. We include separate indicators of initiator and target polity type to pick up linear trends linking regime type and conflict. If democracies initiate or escalate disputes more or less often than do autocracies, the indicator of the initiator’s regime type alone will capture this. Similarly, if democracies
are targets of conflict more or less often than autocracies, the indicator of the target’s regime type will capture this.

We also include indicators of regime similarity to allow for the possibility that there is a lower probability of conflict when both states are highly autocratic. To examine the possibility of either an autocratic or a democratic peace in one step, we compute two interactive variables. We compute an interaction of the initiator’s democracy score multiplied by the target’s democracy score. In this case, both democracy scores range from $-10$ to $+10$, their original ranges. This interactive variable will be high (large and positive) for jointly autocratic and jointly democratic pairs, in a middle range for pairs of mixed regimes (e.g., democracy scores for either state near $0$), and at the lowest value for pairs of dissimilar states (with one large and positive and one large and negative democracy score, for instance). The measure thus captures similarity, and if similar regimes do not fight one another, then this variable will have an effect on the relative risk of conflict while the other democratic peace variables do not. Because we wish to allow for the possibility that there is either only a democratic peace or both a democratic and an autocratic peace, we also include this term squared. This allows for a curvilinear relationship across the range of joint democracy-autocracy (specifically, this allows for an inverted U-shaped relationship, so that jointly democratic and jointly autocratic pairs may have lower conflict probabilities than middle pairs).

Finally, we wish to test whether democracies or autocracies differ in whether they initiate and escalate disputes against similar types of states. We begin by constructing a second measure of regime similarity as $20 - |\text{Dem}_{\text{Initiator}} - \text{Dem}_{\text{Target}}|$. This differenced regime score measure will range from $0$ for similar regimes to $20$ for dissimilar regimes. We subtract this value from $20$ so that similar regimes have a score of $20$ (indicating similarity) while dissimilar regimes have a score of $0$. We then interact the initiator’s regime type with our regime similarity measure. For this interaction, we cannot combine the initiator’s raw democracy score (which ranges from $-10$ to $+10$) with similarity (which ranges from $0$ to $20$). Instead, we add $10$ to that score (so that both range from $0$ to $+20$) so that the interaction is consistently higher for states that are more similar and more democratic. If we allowed the initiator’s democracy score to range from $-10$ to $+10$, then an autocracy times a similar state would have an $X$ value of $-200$, while a democracy times a similar state would have an $X$ value of $+200$, and all dissimilar states would have a value of $0$ from the interaction. A single coefficient on this interaction would not tell us what we wish to know. However, if we transform the initiator’s democracy score to be strictly
positive, then the interaction is also always positive and is consistently higher when a more democratic state is the initiator and a more similar state is the target. Autocratic initiators and dissimilar pairs always have a low value on the variable.

While our specification of democracy is more complex than most found in the literature, it has the advantage of allowing us to test a variety of current and overlapping conjectures relating democracy and conflict. Most analyses in the current literature include dummy variables marking “joint democracy” if both states have a regime score over some level and “joint autocracy” if both scores are under some level, where the level varies widely from study to study (see Bennett 2002 for discussion). States not included in either of those two categories clump together into a third group. Our continuous specification allows for nonlinear effects that vary across the range of joint democracy-autocracy without forcing us to prespecify cutoffs for the joint categories. Overall, we expect disputes among more democratic dyads to escalate to war less often than disputes in less democratic dyads, but we also expect the effects of democracy to differ for different levels of dispute escalation following Senese’s (1997) findings.

8. Expected Utility

Applications of expected utility theory to war initiation and escalation begin with the assumption that we can best model states’ behavior as the result of leaders’ conscious, instrumentally rational choices. In this approach, the anticipated costs and benefits of alternative actions guide leaders’ choices. When faced with a choice between alternatives, rational choice theorists assume that states’ leaders will choose the strategy with the best expected net return—the choice with the highest expected utility. Of the various arguments we investigate, expected utility theories are typically the most carefully developed and come closest to achieving a fully specified theory of war. Based on deductive logic, presentations of expected utility models typically begin with an explicit statement of the model’s assumptions, and, typically, the hypotheses deduced from the model are exquisitely clear. Unfortunately, while theoretically precise and mathematically elegant, such models can also be devilishly difficult to test (Smith 1996a; Morton 1999; Signorino 1999; Lewis and Schultz 2002). Expected utility theorists make very strong claims about the theoretical superiority of their approach, and there are numerous examples in the international relations literature (e.g., Fearon 1994; Bueno de Mesquita et al. 2001; Morrow 1989; Werner 1999).
The empirical tests we present make a substantial contribution to the debate on the relative scientific value of expected utility models, in part because most of the standard criticisms of rational choice models have been theoretical. Attacks on rational choice models come from several directions. Some argue that tests of equilibrium-based models are problematic because a model’s predictions simply reflect the underlying solution concept the authors employ rather than reflecting any particular truth about the nature of politics (Riker 1982; Signorino 2000; Lewis and Schultz 2002). It is more common that critics question whether we should characterize leaders’ decision-making processes as “rational” in the first place (e.g., Lebow 1981; Jervis, Lebow, and Stein 1985; Rosen 2002) or whether the expected utility approach is as comprehensive and encompassing as its proponents claim (Simowitz and Price 1990). Some critics have been willing to make quite strong condemnations of rational choice models. For instance, Green and Shapiro (1994, 6) argue, “The case has yet to be made that these models have advanced our understanding of how politics works in the real world.”

Despite vigorous theoretical criticism, there has been limited critique of the various expected utility theories of war based on systematic empirical analysis. We will investigate most closely Bueno de Mesquita and Lalman’s 1992 domestic variant of the IIG, the most widely cited of these conflict models. We choose to test this particular model for two reasons. First, the model that Bueno de Mesquita and Lalman develop is very ambitious. While most expected utility models depict a relatively narrow puzzle or specific question (see Fearon 1995; Schultz 1999; Powell 2001), the IIG is a game that can apply to almost any international interaction, in that it incorporates both domestic and international political actors in a wide variety of possible outcomes ranging from the status quo to negotiation to war. Of the extant expected utility models, the IIG comes closest to approximating a complete conflict-data-generating process. Second, Bueno de Mesquita and Lalman go to great lengths to develop observable empirical measures for many of the abstract concepts lying at the heart of their model. Most rational choice theorists treat abstract concepts such as similarity of interests, risk aversion, and the costs of war as implicitly immeasurable (Powell 1999).

The IIG is a stylized series of strategic decisions in extensive form that lead to eight possible outcomes for any dyadic interstate relationship (see fig. 4.1). A very general overview is that the game represents an interaction structure between a potential conflict initiator (state A) and a potential target (state B). The game begins with a choice by state A to issue some type of demand to state B. If A does not make a demand, then B may make a demand. If neither state does, then the out-
come of the interaction is the status quo. If one state initiates a crisis by making a demand, then at subsequent decision nodes in the game, actors A and B alternately choose to make further demands (escalating the conflict) or not (resulting in some type of settlement). Each actor makes the choices it believes will yield the greatest expected utility.

An appropriate test of the IIG examines whether its equilibrium predictions match actual dyadic outcomes, as the game predicts a unique outcome for each directed dyad-year. The independent variables for testing the model are not the utility scores. Oneal and Russett’s (1999a) analysis does not represent a thorough test of the model, as the utility scores combine in a nonlinear manner through the structure of the IIG to yield an equilibrium prediction of the game’s outcome. In Bueno de Mesquita and Lalman’s very limited tests (707 European dyad-years), they claimed to find substantial evidence that the predictions of the IIG fit actual conflict behavior. Tests using a larger set of cases seemed to support this claim (Bennett and Stam 2000b, 2000c) but with only a limited set of control variables. Testing the propositions laid out in the IIG is difficult because of the complexity
of generating the data needed to test the theory. This process includes computing utility estimates that combine measures of the similarities of dyadic alliance portfolios, individual national capabilities, dyadic distances, states’ risk propensities, and projected third party choices to join in. Computation of risk propensities in turn involves estimating the hypothetical best and worst security situations in which a state might find itself, a time-consuming computation that involves searching over billions of possible alliance patterns in order to find the best and worst security scenarios for each country-year (see Bennett and Stam 2000 and on-line EUGene documentation).

We begin the tests of the model by estimating the states’ utilities for different outcomes in each directed dyad-year. For each directed dyad-year, we created a prediction of which of the eight outcomes should result. Because these equilibrium predictions are mutually exclusive, we must exclude one from the logit model, the variable predicting the status quo. The MID data and the game do not match up precisely. The MID codes five outcome categories, the game codes eight. For the game outcomes that do not correspond directly to MID outcomes, we need to match them to the most appropriate MID outcome. Since they represent middle levels of escalatory behavior after a demand is made (and a dispute is initiated), we would expect the acquiescence and negotiation equilibria to bear a positive relationship to dispute initiation and escalation to medium levels, but probably a negative relationship to escalation at very high levels. We would also expect a positive association of the war equilibrium with both the initiation of militarized disputes and escalation through all levels of force.

9. Geographic Contiguity

While there is no “theory of geographic contiguity” per se, some argue that contiguity is a necessary condition for war between most pairs of states (Bueno de Mesquita 1981; Maoz and Russett 1993). Maoz and Russett argue that, for all but the most powerful states, the challenge of projecting military power across long distances presents a hurdle that most states simply cannot clear. The vast majority of states simply cannot fight one another across long distances, as they lack the power projection and logistical capabilities needed to do so. As a result, distant pairs of states that might otherwise consider the mutual use of force remain eternally quiescent. In addition, contiguous states may be more likely to have unresolved disagreements over the settlement of territorial issues between them that noncontiguous states do not, either because
noncontiguous states have no point of contact or because the salience of bordering territory is higher than more remote territory (Vasquez 1996). Others have argued that contiguity is merely incidental to conflict. Holsti (1991), for example, argues that in fact only certain strategic borders are friction prone. Empirically, however, studies have continued to find that geographic contiguity has a powerful positive association with violent conflict (Diehl 1991; Bremer 1992; Kocs 1995).10

Although contiguity is one of the selection criteria for politically relevant dyads, its inclusion in our analysis of that sample is not a problem, as the variable serves to distinguish between contiguous dyads and those that are noncontiguous but where one state is a major power. We measure contiguity as a dummy variable, with a “1” marking a dyad in which the states are contiguous on land and a “0” otherwise, based on the COW 1993 contiguity data set.

10. Nuclear Deterrence

Many scholars have argued that nuclear deterrence—particularly MAD—between the superpowers was a key factor in avoiding war between the major powers after World War II (Glaser 1992; for a contrary view, see Jervis, Lebow, and Stein 1985). Because each state knew that it faced unavoidable destruction—or, rather more likely, unacceptable damage, in the case of a nuclear exchange—neither side would rationally take the final escalatory step of initiating a nuclear attack. The deterrent effects from a threatened nuclear exchange on lower levels of conflict were less clear, though. Some argued that nuclear armed states would be very careful about taking lower-level aggressive steps that might escalate conflicts to a level at which there was a risk of nuclear exchange. However, others argued that the certainty that neither side would use its nuclear weapons made the world safe for conventional confrontation between the superpowers, leading to a need for large conventional defensive forces as well as massive nuclear arsenals. The evidence cited for the effectiveness of nuclear deterrence appears in the lack of a war between the United States and the Soviet Union during the cold war. However, the nuclear deterrence explanation is almost perfectly collinear with other equally compelling explanations, including post–World War II war weariness, bipolarity, and increasing interdependence (see, e.g., Mueller 1994). The evidence in favor of the pacifying effects of superpower MAD is limited at best (specifically, the single observation of no superpower war).

As nuclear weapons spread beyond the superpowers, more testing of
their effects is now possible. Huth and Russett (1984) provide some additional evidence, finding that nuclear weapons appear to have some deterrent effect in crises between the superpowers and other states. The belief in the efficacy of nuclear deterrence has led some to suggest, in fact, that we should actually encourage the spread of nuclear weapons. For instance, Waltz (1995, 4) argues, “the gradual spread of nuclear weapons is better than either no spread or rapid spread. . . . Nuclear weapons make wars hard to start. These statements hold for small as for big nuclear powers. Because they do, the gradual spread of nuclear weapons is more to be welcomed than feared” (see also Bueno de Mesquita and Riker 1982).

The past three decades have seen an increasing number of dyads where one or both states have nuclear capability. On one hand, if nuclear weapons affect states’ calculations of the likely costs of conflict, and if nuclear weapons may actually be used, then we should see the deterrent effect of such weapons in any dyad where they are present. On the other hand, if there is a nuclear taboo that prevents states from seriously considering the use of these weapons, then they may have no effect (Tannenwald 2001). Empirically, our investigation seeks to discover whether there is now enough variation in the possession of nuclear weapons across states and time to sort out their effects from other factors.

We include three dummy variables concerning the possession of nuclear weapons. The first marks whether the potential conflict initiator in a dyad has nuclear capability, the second marks whether the potential target has nuclear capability, and the third marks whether both states have nuclear capabilities. The third variable is particularly important to test arguments that nuclear powers are unlikely to use nuclear weapons against nonnuclear states for fear of adversely affecting their reputations. With these variables, we can see whether possessing a nuclear capability affects a state’s behavior toward nuclear armed targets or nonnuclear targets. We code the following states as having nuclear capability: the United States from 1945; the USSR from 1949; Britain from 1952; France from 1960; China from 1964; Israel from 1973; India from 1974; Pakistan from 1986; South Africa between 1980 and 1991; and Ukraine, Kazakhstan, and Belarus in 1992.

11. Power Transition

In stark contrast to balance of power advocates, Organski and Kugler (1980) argue that periods of power parity are more likely to lead to war
than periods of power preponderance. Unlike simple power preponderance arguments, however, power transition logic suggests that it is power parity plus shifts in power, plus dissatisfaction with the status quo on the part of the overtaking state, that lead to conflict. Power transition theorists maintain that it is when states are passing one another in relative power (with changes driven by domestic political and economic development) that the probability of war and conflict is greatest. Organski (1958) argues that states gain benefits from a higher position in the international hierarchy of states, in part due to subjective notions of status and in part due to greater coercive capacity, and so constantly strive to rise in the system and become the system’s leader. However, Organski assumed that the system leader would not willingly concede its leadership role when challenged, particularly if the leader and challenger differ in their preferences for the system’s rules and distribution of material resources. The resulting competition for the system’s leadership role increases the chances for violent conflict. Power transition arguments specifically focus on the decisions made by state leaders, with Lemke and Kugler (1996, 4) explaining that

The cornerstone of power transition theory is that parity is the necessary condition for major war. However (unlike balance of power theories), power transition theory does not deny the importance of decision-making, nor does it imply that structures determine outcomes. Rather, power transition theory contends that parity sets the stage where decision makers can, but need not necessarily, choose major war as a viable alternative. [italics in original]

The definition of what constitutes a “major war” is an important distinction. Organski and Kugler focused on what they called “system altering” wars, which tend to be more easily identified than creating distinctions between wars fought on a smaller scale. (For an exception to the claim that system altering or major wars need to be materially large, see Bueno de Mesquita 1990). The problem with following Organski and Kugler’s focus on system altering wars is threefold. First, it is unclear what constitutes an alteration of sufficient magnitude for a war to be “major.” Second, these wars, as they inductively describe them, are so rare that their claims are untestable with the means employed here. Third, “major wars” can only be identified ex post, creating a set of epistemological problems about studying the origins of conflict.

Power transition theorists initially focused on transitions between the dominant state in the international system and close contenders as those that are most likely to produce conflict. Organski and Kugler, for
example, examined relations only between the most powerful state in the system and those that they reasonably expected to challenge it, operationalized as states within 80 percent of the aggregate power of the dominant state. From this original version, power transition theorists have developed the logic in several ways. First, some now include a focus on transitions within regional hierarchies. For example, Lemke (2002) finds that power transitions do occur in this setting. Second, others have focused on transitions within rival dyads. Huth, Bennett, and Gelpi (1992) found that power transitions among all great power rival dyads increased the initiation of militarized disputes. Third, others focus on how satisfaction and power may function in driving conflict within transitions (see Lemke and Werner 1996; Lemke and Reed 1998; Oneal, De Soysa, and Park 1998). Others have examined the precise timing of the parity, power shift, and war relationship to try to determine which moment is the most dangerous. It might be just before parity emerges or the period immediately after the rising state overtakes the declining state.

Lemke (2002), who advocates the expansion of power transition logic beyond its original scope to include a far greater population of dyads, suggests that the narrow initial focus of the approach is inappropriate. While he focuses on regional leaders and regional hierarchies, an extension of his argument suggests that we should examine every dyadic relationship for a dominant state, even if neither state is the top state in the international system. Following this line of argument, we might expect to see power transition dynamics apply in any dyad we select. This more inclusive power transition logic argues that we should expect to observe dyadic parity and power shifts associated with greater incidence of conflict, the direct opposite of balance of power arguments (in its dyadic variant). For dyadic-level power transition logic to make sense, the key distinction comes down to what we believe a system of states to be. It might be the entire system of all states, a regional subsystem, or simply a system of two. Physical geography may outline the boundaries of the various systems we might expect to find. We begin our empirical investigation as the broadest possible notion of systems, at the dyadic level.

Our measures of power transition here include measures to examine the relationship between power parity, dynamic power shifts, satisfaction with the status quo, systemic leadership, and the incidence of war linked together through power transition logic. First, our basic balance of power variable (discussed previously) captures the effects of static parity. The sign for the dyadic power balance parameter should be positive from the perspective of power transition advocates and negative
for balance of power advocates. Second, we compute a variable measuring dynamic dyadic transitions in power as an interaction of equal power and differences in growth rates, not taking into account the satisfaction of the actors that may be passing one another. This variable results from the interaction of two variables measuring “equality” and “growth,” producing a coding of how intense the power transition is in every dyad in our study. Since power transitions occur only when states are close to each other in power, we begin by measuring how close to power parity the two states are. We start with our continuous balance of forces measure, which ranges from 0.5 when the states are equal in power to 1 when they are unequal. We then transform the variable into the range from 0 when unequal to 1 when equal by computing degree of equality = $2 - (2 \times \text{balance of forces})$.

We then compute a five-year moving average of the two states’ annual percentage growth, using the COW national capabilities index (CINC). We then difference these two rates and take the absolute value to obtain a measure of how quickly the gap between the two states was either shrinking or growing. For example, if one state had average growth of 2 percent and the other average growth of 6 percent, the difference was 4 percentage points. According to power transition advocates, this situation of rapid transition should be more conflict prone than a situation where the difference was only 1 or 2 percentage points. Here the intuition is that the states should have greater difficulty designing diplomatic solutions, thereby becoming more prone to miscalculations and strategic error during times of rapid change or greater instability. We then create an interaction term with the degree of equality and the difference in growth rates. The value of the variable is high when the two states are close in power and the differences in their rates of growth are large. In contrast, the value of the variable is low when the states are far apart in terms of power or have similar rates of growth. Power transition theorists would expect the variable to correlate positively with conflict behavior.

Third, to capture Organski’s (1958) original notion of hierarchy and status, or more recent variants’ emphasis on relative satisfaction, we follow Lemke and Reed’s (1998) arguments and compute a variable that also incorporates states’ satisfaction with each other’s international policies. We measure the satisfaction of two states with one another by using the tau_b score of the two states’ alliance portfolios. Tau_b is a correlation coefficient representing the similarity of the two states’ alliance portfolios. Lemke and Reed argue that states that are dissatisfied with the nature of the dyadic hierarchy will seek different allies, creating a measure of revealed preferences. States satisfied with their status or role
in the hierarchy should have relatively similar alliance portfolios. Note
that this logic is inconsistent with Walt’s (1984) balancing and band-
wagoning logic. Walt argues that it is the perception of threat and not
satisfaction or dissatisfaction that drives a state’s alliance choices.

We then multiply the tau measure by $-1$ so higher values represent
disagreement and interact it with our dyadic power transition measure
to obtain the final variable. Finally, to contrast Lemke’s logic with Or-
ganski’s original argument, we compare power transitions in all dyads
to the effects of power transitions among the most powerful states in
the system by creating a final interaction of the dyadic transition mea-
ure. Following Organski’s argument that Britain was the system leader
through World War II and the United States after that, we interact our
transition variable (measured with satisfaction) with a dummy variable
marking (1) dyads through 1945 where Britain is one dyad member and
(2) dyads from 1946 on where the United States is one member. These
dyads could be involved in a system leadership transition. The measure
will have a value of zero for all dyads not involving the system leader,
and so the effect of the variable on conflict will always be zero in these
cases. Other values will allow us to estimate whether there is any addi-
tional probability of conflict in situations involving the dominant state.

12. Trade Interdependence

The argument that increased international trade lowers conflict be-
tween nations dates back centuries. The idea was particularly in vogue
in the period before World War I, but the nature and length of that war
helped to discredit the liberal idea that interstate trade would make war
obsolete (Carr 1946). More recently, the argument that trade should
lower the likelihood of armed conflict between states has been revived
in arguments about the so-called Kantian peace (Russett and Oneal
2001). In Triangulating Peace, Russett and Oneal argue that an inter-
action of democratic institutions, interstate trade, and international or-
ganizations contributes powerfully to international cooperation and the
reduction of international conflict. Their straightforward argument is
that, when states are interdependent, they have incentives to avoid con-
flicts that may result in a costly disruption of trade. The more one state
depends on another for trade, the greater the likely cost of adjusting to
a reduction or cutoff of trade flows. Trading countries may also learn
about one another, lowering conflicts of interest and misconceptions
about one another. Oneal and Russett operationalize trade interde-
pendence as a contributing factor in the reduction in conflict by com-
puting the mutual trade dependence of each state in a dyad by dividing total dyadic trade by GDP. For each dyad-year, this gives us two trade dependence measures, one for each state.

Neither the arguments, the operational methods they employ, nor the data used by Oneal and Russett are without controversy. In terms of the trade and GDP data, which are notorious for their unreliability, particularly in the early years of the data series, Oneal and Russett interpolate between observations to reduce missing data. It is important to note that in later years of the series, where they cannot reliably fill in gaps in the data, they recode International Monetary Fund (IMF) data originally labeled “missing” as zero trade. This is a particularly problematic issue regardless of one’s solution. By dropping cases rather than assuming the trade to be zero, the loss of cases may introduce substantial selection bias into analysis. Alternatively, coding the data as zero may introduce substantial nonrandom measurement error if the two states actually had any trade. GDP data, which does not exist directly for most countries before World War II, comes from the estimates of economic historians, with key assumptions made to convert estimated relative growth measures to comparable dollar values.

Other scholars have used alternative measures, believing that dyadic trade divided by GDP is a poor measure of trade dependence (Barbieri 1996). In terms of the argument linking greater trade dependence to more quiescent international politics, it may be too simplistic to argue that trade provides incentives for states to avoid conflict. A state dependent on another may have incentives to provoke conflict if such conflict might improve terms of trade. Canada and the United States have been involved in trade disputes multiple times over the past forty years that escalated to the show of naval military force. By using their naval forces during a trade dispute, smaller countries such as Canada and Spain are able to demonstrate how seriously they take the trade dispute, which may be far more salient for them than it is for the much larger United States. Low-level conflict with a trading partner may also provide a means of signaling dissatisfaction with the relationship. Moreover, the cutoff of dyadic trade may not always lead to significant costs for either side of the dispute. Firms typically use multiple suppliers. The development of a multinational free-trade system means that states have ready alternative suppliers willing and able to substitute for trade lost during a low-level militarized dispute between two countries. This ability to substitute for lost trade with one partner lowers the economic costs of conflict associated with trade disruption (Brooks 2001).

Finally, different types of trade, or trade in different types of goods with varying elasticity of demand, may be differentially associated with the
economic cost of conflict (Crescenzi 2000). Raw materials are more substitutable than high-tech imports, for instance. This would suggest that total dyadic trade is, at best, a crude indicator of the likely cost of trade disruption.

Setting aside criticisms of measures and theory for the moment, we include two trade dependence variables in our analysis, one measuring dyadic trade as a proportion of GDP for the potential initiator and the second measuring dyadic trade as a proportion of GDP for the potential target. Prior work on trade (e.g., Oneal and Russett 1996) typically combines these variables and uses the minimum of the two trade levels in the analysis. This is largely because previous studies focused on nondirected dyad data, following the so-called weak link argument (Dixon 1993). As noted in chapter 3, in a nondirected dyadic analysis, we must transform the two variables somehow to remove the directionality of the initiator and target from the variables. The weak link assumption that accomplishes this argues that it is the less constrained of the two states in a possible conflict that will drive any conflict initiation and escalation. If one state in a dyad is highly trade dependent on the other, and therefore cannot risk losing this income from trade, it is likely (according to the weak link argument) to be more constrained than the less trade-dependent partner in the dyad. If the second state is largely independent of the first, as is often the case when great powers interact with small states, and hence trade is less salient than it is for the first state in the dyad, the second state will ignore the trade constraint on the first. Trade between the two may then appear to have no restraining effect if the second state is the aggressor. We include the measures of trade dependence separately for two reasons. First, because we have a directed dyadic design, we can easily include both measures. Second, we find the logic of the weak link argument applied to trade to be unsatisfying.

Even with the interpolations and data assumptions made by Oneal and Russett, there are substantially more missing data on trade dependence than is the case with any other variable in our data set. Because the trade data are not missing randomly, we are particularly concerned about how including trade may bias our other results (closed autocratic states do not report accurate GDP and trade data, and a large proportion of states have missing data before World War II). When we add the trade variables to our analysis, we lose 23 percent of the cases in our analysis of all dyads and 38 percent of the cases in our sample of politically relevant dyads. Such a major loss of data, particularly when it is systematic and correlated with other factors we wish to analyze, will likely bias our results on other variables. As a result, we
add trade to our model in a secondary analysis after we analyze the larger, more complete data set.

INTERNATIONAL SYSTEM LEVEL OF ANALYSIS

The final area we investigate is the nature of the international system and its effects on the behavior of nation-states. Systemic theories of international conflict have a long history in the study of international politics. Realists traditionally start the discussion by noting what they refer to as the “self-help,” or anarchic nature, of the system of states. Indeed, for many scholars, the defining characteristic or key feature that distinguishes international politics from any other variety thereof is the lack of central authority or binding contract enforcement mechanisms. The absence of a recognized central government with its own police power to enforce international law (and the anarchy this implies) leads some to then conclude that the distribution of power and beliefs about system-level power are where we should look to understand the behavior of states (Waltz 1979; Wendt 1999; Singer, Bremer, and Stuckey 1972; Morgenthau 1956). Perhaps the leading system-level scholar of the past fifty years, Kenneth Waltz argues powerfully that any theory or explanation of international relations that looks to levels below the international system level is no longer a theory of “international politics” but instead is merely a theory of foreign policy—implicitly, somehow less worthy of study. Waltz’s argument begs the question of what the politics of nations are, however, and what determines the particular relations of states in the international system as compared to general levels of conflict throughout the system.

We take a somewhat different tack concerning system effects and their role in international politics. In our view, since international politics results from the interaction of states’ individual and collective foreign policies, and the interaction of decisions and acts of foreign policy, there is no relevant distinction to make between foreign policy and international politics. This does not imply that we can safely ignore the effects of varying systemic characteristics. Rather, it suggests to us that to analyze the nature of international politics with a myopic focus on system structure, without reference to national decisions about foreign policy, would be folly. In the section that follows, we summarize the dominant and directly testable propositions or conjectures made by the leading international system-level theorists. By focusing on those that are directly testable, here we necessarily restrict ourselves to theories or arguments that focus on the material distribution of capabilities. Realist
systemic explanations tend to focus on such capabilities, although Van Evera (1999) develops a notable exception. For an alternative view, based almost entirely on the subjective beliefs of the actors (which we cannot test directly), see Wendt 1999. To understand better the nature of the shifting beliefs to which Wendt and other constructivists refer but fail to specify in rigorous or testable fashion, see Savage 1954 and Converse 1964. We begin our discussion of system-level effects with the long cycle literature.

13. Economic Long Cycles/Kondratieff Waves

Some economists have argued that the world economy goes through long-term cycles of growth and stagnation (Kondratieff 1935). Building on the assumption that states need significant financial resources to prosecute interstate war, Joshua Goldstein speculates that there should be strong correlations between the international system’s economic growth cycles and the relative likelihood of interstate war. He developed arguments about the effects of these economic cycles on international conflict through two mechanisms (Goldstein 1988, 1991; for a critique, see Beck 1991). First, as an exceedingly expensive activity, Goldstein argues that war requires economic prosperity, which is most common throughout the international system during periods of economic growth. Second, periods of systemic economic growth may also lead to heightened competition for world resources and markets. The ability (through prosperity) and demand (through competition) to fight for territory or resources may result in more and/or more severe wars during periods of prosperity, particularly among the dominant economies in the international system. Notably, Goldstein does not develop a detailed theory of war initiation. Rather than carefully developing a model of the causal path to war during periods of growth compared to periods of relative scarcity, he devotes the bulk of his work to developing the empirical basis for his claim that systemic economic cycles actually exist in the first place.

Goldstein found substantial evidence of long waves both in economic conditions and in wars but found less consistent evidence of a relationship between them. In particular, Goldstein found that, while major power wars were no more frequent during economic upswing phases than during other phases, they were more severe. Here we do not attempt to develop the missing theory that might help explain why several of Goldstein’s hypotheses failed to work out. Instead, we aim to test more rigorously the empirical proposition that there are links between the system’s economic growth cycle and interstate conflict.
We use Goldstein’s (1988, 247) coding rules for periods of economic upswings and account for two representations of the “economic upswing” variable beyond 1975 when Goldstein’s coding ends. His coding rules are somewhat problematic because he argues that there may be cycles within cycles. For instance, Goldstein initially codes 1968–75 as an economic downswing, which we project forward leading to a coding for 1968–92 as a global economic downswing. However, Goldstein also seems to suggest that the longer period 1940–80 is really a larger economic upswing (218). Therefore, we also test a coding rule of labeling 1940–92 as one long upswing. We found similar results regardless of which coding rule we employed. In our final analysis presented here, we include a dummy variable that distinguishes those years that were in an economic upswing (1) from those that were not (0) using the first of these two coding methods.

14. Hegemonic Stability

Much like the logic behind the various power transition arguments, the literature on the role that hegemony plays in international politics focuses on the pattern of changing capabilities among the top states in the international system. Hegemonic stability models focus on the effects of the presence or absence of a hegemon in international affairs (Kindleberger 1981; Gilpin 1981; Krasner 1976). Hegemonic stability scholars argue that the international system is most stable when one state controls a preponderance of the international system’s resources and can impose its policies upon other members of the system. Periods of bipolarity or multipolarity, where there are two or more great power states of relatively similar size competing for dominance in the system, will be less stable than periods of hegemony. Hegemons are unable to maintain their position at the top of the international hierarchy forever, however (Kennedy 1987). Varying growth rates and/or the costs of system leadership lead to inevitable cycles of rise and decline, and conflict will eventually occur when the single great power is no longer able to ride herd on the other states in the system. From a related cyclic perspective on “power cycles,” Modelski (1987) similarly argues that a cycle of power and wealth driven by the disparity between the costs and benefits of monopoly in the international system leads inevitably to the successive rise and decline of world powers, with periods of dominance being the periods of greatest stability.

Various authors have measured hegemony in a variety of ways, emphasizing different aspects of the phenomenon (for example, by
measuring overall military resources, economic power, or naval power). We employ Gilpin’s (1981) identification of the two periods of hegemonic governance in the international system relevant for our analysis. According to Gilpin, from 1816 to 1918 the system was under British hegemony, and from 1946 through the end of our study the United States was a hegemon. We include a variable marking the period of British hegemony to 1918; the post-1945 period of U.S. hegemony is marked in our analysis by the same variable marking the bipolar system. Because Gilpin codes the postwar period as hegemonic, we create a problem of collinearity in that Waltz and other scholars consider the same period to be bipolar.

With these characterizations of hegemonic versus multipolar periods or systems, our goal is to establish whether the postwar period has a different baseline risk of conflict compared to the interwar period and pre–World War I. For the moment, we ignore the inferential problems that Gilpin’s and Waltz’s dissimilar codings of the postwar period create. Note that the empirical prediction of these theories is that long time periods should be different from one another because of hegemony (bipolarity) or the lack thereof. This is a rather unspecific prediction, and there could be other characteristics of the system over long periods that make it more or less dangerous. The test is essentially, while controlling for other factors, whether the average baseline rates of conflict over these periods differ. This admittedly unsatisfying and overly broad test is where these loosely specified arguments leave us with regard to expectations about the incidence of interstate conflict and war. Hegemonic stability models typically predict that both the pre–World War I and post–World War II periods should be more stable than the interwar period, when there was not a clear hegemon.

15. International System Polarity

Waltz (1979) and other neorealists made three major modifications to the classical balance of power approach. First, Waltz argues that states do not pursue power directly, but rather they seek to maximize their security. This subtle difference (subtle because power is an essential factor determining a state’s national security) allows for the possibility that states will not constantly seek to expand nor to always choose guns over butter. Second, Waltz and other neorealists argue that polarity is the key attribute of the international system that critically affects states’ pursuit of security. Third, Waltz conjectures that balancing—states acting to oppose potential hegemons—is automatic.
For neorealists, polarity or the number of large great powers present in the international system is of prime theoretical and empirical importance. In a bipolar system, where there are two major powers clearly separated from the other states in terms of relative capabilities, the major powers can achieve security through internal balancing. They do not have to rely on alliance partners who may prove unreliable or drag them into unwanted conflicts (Snyder 1984). In a multipolar world, though, the capabilities of several states simultaneously affect the security of any single state, forcing all states to mind the actions of all others and to pay much greater attention to alliances. The uncertainty inherent in this complex system may lead to a greater incidence of miscalculation, in turn leading to conflicts that would have been avoidable under the more transparent conditions of bipolarity. The evidence supporting this argument essentially boils down to the observation that there was no war between the United States and the Soviet Union during the forty-five years of the cold war.

Critiques of neorealist style arguments about the effects of system polarity come from several directions. The first wave of logical counter-arguments came from Deutsch and Singer (1964), who argue that states would in fact be less likely to attack others under multipolar conditions, first, because uncertainty forced states to divide their attention among many potential foes and, second, because cross-cutting ties could exist between multiple states that would serve to reduce hostility. Divided attention would make it less likely that any state could be confident enough about winning that it would start a conflict that might inadvertently escalate into an unwanted war; cross-cutting ties would reduce overall hostility levels and provide positive incentives to avoid conflict. Wayman (1984) in turn argues that there were different kinds of polarity, what he termed “power polarity” and “cluster polarity,” with the more nuanced characterization seeming to reconcile the competing arguments. Using finer measures of polarity, Wayman found evidence that bipolar systems were somewhat more peaceful than multipolar ones.

Criticizing polarity arguments from a different tack, Bueno de Mesquita (1978, 1981) argues that these simplistic system-level arguments about polarity were inadequate because they implicitly assumed that there is a dominant systemic risk attitude. According to Bueno de Mesquita’s argument, Waltz implicitly assumes risk-acceptant decision makers who react to multipolar uncertainty by starting conflicts, while Deutsch and Singer implicitly assume risk-averse decision makers who avoid conflict under conditions of uncertainty. The reconciliation of the two perspectives plus concern for risk attitude required scholars to
resort to a lower level of analysis, namely, the decision level. Huth, Bennett, and Gelpi (1992) and Huth, Gelpi, and Bennett (1993) tested some of these arguments about the interactive nature of polarity and risk-taking propensity and found that the interaction of risk attitudes and system structure did help to account for a larger proportion of international conflict initiation, but not conflict escalation. However, they also found that, when tested statistically, a single variable marking bipolarity did not correlate with conflict initiation.

Other evidence against Waltz's argument appears if we look at the length of time the system went without a war under bipolarity. When the system was bipolar, forty-five years passed (1946 to 1991) without a war between the poles. When it was multipolar, there were thirty-four years without a major power war between 1870 and 1904, thirty-eight years without a major power war between 1816 and 1854, and forty-four years without a war in Europe between 1870 and 1914. Critics of system structure stories argue that these numbers show that the pacifying effects of bipolarity have been overstated, especially given the much larger number of opportunities for war between major powers before 1945 (Hopf 1991).

Regarding the initiation and escalation of disputes among major powers, as well as the incidence of disputes among nonmajor powers, note that one of the reasons that a bipolar system is stable, according to Waltz, is that the two major powers will be vigilant and will respond quickly to any challenge by the other major power. Because they will respond quickly and because both states are resolved, war should not occur. Nevertheless, even if this logic is correct, this does not necessarily imply that there will be fewer disputes between the major powers in such a system. In fact, because they are acting to counter each other's moves quite quickly, there might be more disputes short of war than in other types of systems. If this were the case, then we would expect minimal reduction (or even an increase) in the risk of dispute initiation between major powers during bipolar periods and a negative effect only for dispute escalation. Extending neorealism logic to minor powers, the major powers may enlist the aid of proxies, whom they might encourage to fight (or resist) the other major power, leading to a further increase in disputes. According to this logic, there may actually be more disputes among minor power states during bipolar periods than during multipolar eras.

Again, as was the case with our tests of the democratic peace proposition, here we do not test the mechanisms claimed to lie behind the various polarity-based arguments and conjectures found in the litera-
ture. Instead, we more carefully test whether the post–World War II pe-
period was more or less prone to conflict than we would expect, given the
effects of the various other factors believed to affect the likelihood of
conflict. As mentioned previously, using the periods as delineated by
Waltz and Gilpin, the post-1945 era is either bipolar or hegemonic, de-
pending on whose classification one chooses. Our dummy variable
marking hegemony will measure effects on conflict behavior common
to both 1816–1918 and 1945, while the post-1945 variable will meas-
ure effects from 1945 on. If the post-1945 era is more peaceful than
other periods, however, there is no way to tell whether the cause is
hegemony or bipolarity (Gilpin’s and Waltz’s variables have no further
variation). As a result, when we present our results we will discuss time
periods rather than focusing on hegemony or bipolarity. We create a
dummy variable to mark bipolarity, coding the variable as a “1” when
the year of possible dispute was after 1945 and a “0” otherwise. We
also created an additional variable by combining the post-1945 vari-
able with one marking the U.S.-USSR dyad. That is, the variable was
coded a “1” for dyad-years after 1945 between the United States and
the USSR and a “0” for other dyad-years. According to Waltz’s argu-
ment, this variable in particular should have a negative coefficient, al-
though the coefficient on bipolarity alone might have a positive coeffi-
cient without violating the model’s logic. Because of collinearity, we
were unable to test this variant and so, confronting an early limit to
knowledge, include only the broader measure, which differentiates all
post–World War II dyads from others.

16. Systemic Power Concentration and Movement

Reacting to what they saw as flaws in the many arguments and meas-
ures relating polarity and power concentration in the international sys-
tem to the incidence of conflict, Singer, Bremer, and Stuckey (1972) de-
veloped a set of sophisticated quantitative measures for various aspects
of system structure, including a continuous measure of systemic power
concentration that Mansfield (1994) recently reanalyzed. Following
standard “realist” logic, Singer, Bremer, and Stuckey reiterated the off-
cited claim that the systemic distribution of military-industrial capa-
bilities is associated with important effects on war and peace. Unfor-
tunately, the precise causal mechanisms linking these systemic
characteristics to state leaders’ decision to go to war remain unclear in
much of the literature on this topic. Singer, Bremer, and Stuckey argue
The concentration of power in the system should correlate with the uncertainty that leaders face when confronting international crises. The theoretical arguments linking uncertainty to war are not well specified. They require as yet missing linkages to state-level variables such as domestic political institutions and to individual-level variables such as risk attitude. Here we aim to clarify the empirical relationship between measures of system structure and the incidence of war and actual conflict behavior.

Both the earlier Singer, Bremer, and Stuckey study and Mansfield’s recent work included three variables of capability concentration, change in concentration over time, and the temporal movement in power shares between states. Mansfield added the square of concentration to account for a potentially nonlinear fit between capability concentration and the onset of war. Despite using similar measures, the findings of these studies point to quite different conclusions. Singer, Bremer, and Stuckey found that power concentration and increases in this concentration have a negative association with annual nation-months of war for the entire period 1816–1965 but that movement in power share from one state to another correlates positively with war (although their findings reverse for the nineteenth century). After controlling for the related effects of hegemony and polarity, Mansfield found strong support for a positive relationship between concentration and war but a negative relationship between concentration squared and war that suggests an inverted U-shaped relationship between power concentration and war. He also found little consistent support for any particular relationship between power movement, changes in power concentration, and war. Unfortunately, Mansfield’s results were inconsistent across different operational measures of system structure.

In our analysis, we include the three main variables of systemic power concentration, change in concentration, and power movement. We employ the “concentration” and “movement” measures for the interstate system defined in Singer, Bremer, and Stuckey 1972. This entails measuring each variable annually and then computing a five-year moving average. We recomputed the power concentration variables using the most recent COW data, and we use all of the states in the system for our coding. “Change in concentration” averages the annual change over the past five years. The value of the concentration index is 0 at an equal distribution of capabilities in the system and 1 when one state holds all capabilities; the related change in concentration variable is high when concentration increases. The value of the power movement index takes higher values when there have been more (and larger) capability-share shifts between states.
This brings us to the end of our list of models and conjectures that we will test in the next chapter. In summary, in the analysis we will examine the following:

**State Level of Analysis**
1. Democratization (six variables)
2. Polity Change and Externalization of Violence (two variables)

**Dyadic Level of Analysis**
3. Alliance and Defense Pact Membership (two variables)
4. Arms Races (one variable)
5. Balance of Power (one variable)
6. Conventional Deterrence (one variable)
7. Democratic Peace (five variables)
8. Expected Utility (four variables)
9. Geographic Contiguity (one variable)
10. Nuclear Deterrence (three variables)
11. Power Transition (three variables)
12. Trade Interdependence (two variables)

**International System Level of Analysis**
13. Economic Long Cycles/Kondratieff Waves (one variable)
14. Hegemony Stability (one variable)
15. International System Polarity (one variable)
16. Systemic Power Concentration and Movement (three variables)

Importantly, while these sixteen arguments represent just a sample of the numerous explanations of the path to war, they require thirty-seven variables to test them. Keep this in mind as we begin to confront some very real limits to our ability to make arguments about the origins of war in general as opposed to the particularistic origins of individual wars. If we find statistical support for just a few of these variables, we will be on safer ground claiming that there are a relatively small number of systematic causal factors leading to war. However, if we find that the majority of the factors identified in this chapter relate systematically to the incidence of war, confidence in our ability to generalize erodes substantially. Stating the puzzle in a somewhat different way, we next begin to tackle the question of whether the greatest common factor leading to
war is that most wars have little in common with those preceding them. In the next chapter, we begin by testing to see if there is a statistically significant relationship between the variables associated with each argument and the presence and level of conflict behavior in each dyad. We then look more closely at the substantive effects associated with carefully controlled changes in the various operational measures.