7 CONCLUSION

life’s but a walking shadow, a poor player
that struts and frets his hour upon the stage
and then is heard no more: it is a tale
told by an idiot, full of sound and fury,
signifying nothing.
—Shakespeare, Macbeth

The analysis in the previous chapters fits into a research mosaic that has been developing over the past sixty years. The quantitative study of the origins and escalation of violent conflict between nations now has a history spanning nearly three-quarters of a century. Published behavioral works on the origins of war trace immediately back to a series of papers published in the 1930s by Lewis Richardson in the British science journal Nature and to Quincy Wright’s A Study of War (1942). Both Richardson and Wright firmly believed that in order to eliminate the pox of war we would need to understand it better, both from a theoretical perspective and, as has largely been the case here, from an empirical one. Their work laid the foundation for several generations of scholars whose best efforts went to building and testing explanations of the onset of war, many of whose work we summarize here. Some have had greater success than others. Progress in social science, as in any science, comes from taking steps forward and backward—steps of building new ideas, refining old ones, and pausing for retrospection and reflection.

Among Richardson’s (1960a, 1960b) most important conclusions based on his data was that chance, and not systematically identifiable causal relations, regulates the distribution of war with respect to both its beginning and its ending. The notion that chance, or the stochastic component of the war-generating process, plays a powerful force in determining war’s onset thus has a long history. Historically, this has typ-
ically been an unsatisfying explanation. Consistent with John Donne’s weary claim about all men that “Thou art slave to fate, chance, kings, and desperate men,” Richardson’s conclusion remains for many an unhappy one, as he began the process of discovering the limits to our knowledge about the onset of war. While Richardson was able to provide powerful descriptions of the wars that occurred in the past, he largely failed in coming to grips with the fundamental question that motivates many of those who study the deadly quarrels between nation-states, namely, what are the systematic causes of war?

The next generation of social scientists working on this puzzle greatly broadened the search for theoretical explanations of war. While doing so, they developed increasingly sophisticated statistical tests of their propositions. Current social scientists working in the empirical tradition owe a huge debt to J. David Singer, whose greatest contribution may have been to democratize the statistical study of war through the development and distribution of detailed data on many of the possible correlates of war (Singer and Small 1972, 1982, 1993). As we take our place at the contemporary end of this research program, we recognize that the questions we have posed and the answers we have provided differ more in manner of degree than in kind from those that precede this work.

If there were one single story to take from this book, it would be that there is no single story of war. In many ways, we are as uncertain about the causes and likely timing of any individual war today as we were in 1942 when Wright initially published his study of war. As our degree of certainty about the inherent limits to knowledge on the origins of war sharpens, we can see the proverbial glass as either half-full or half-empty—we can choose to be optimistic or pessimistic about the current and future state of knowledge and the nature of the unknowable.

As we move to discuss our general findings and the uncertainty associated with our attempts to evaluate the myriad arguments and empirical conjectures about the causes of war, it would be wise to keep in mind the following Socratic dialog that opens Richardson’s *Arms and Insecurity* (1960b, overleaf):

*Politicus:* What are you trying to prove?
*Researcher:* In social affairs it is immoral to try to prove.
*Fidor:* Yes. One should have faith that God will provide. He that cometh to God must believe that He is and that He is a rewarder of them that diligently seek Him.
*Researcher:* I meant that in social affairs where proof is seldom rigid, and where prejudice so easily misleads, it is best not to
start with a fixed opinion. He that comes to research must be in doubt, and must humble himself before the facts, earnestly desiring to know what they are, and what they signify.

In the remainder of this chapter, suitably humbled by the facts as best we know them, we turn to a discussion of what they signify.

A CONSERVATIVE BUT OPTIMISTIC VIEW

One way to judge progress in any scientific endeavor is to pose the question “Which hypotheses perform best?” In this analysis, we have carried out a broader comparative test of many major international conflict hypotheses and conjectures using a larger data set than executed previously, and in doing so we have integrated information simultaneously across several levels of analysis, an impossible task using a qualitative approach. Analyzing the large number of observations and variables has given us many more degrees of freedom and a much improved ability to make carefully controlled comparisons than would be the case had we used other techniques and, correspondingly, more protection against spuriously attributing causal power to otherwise random correlations in the empirical results. The resulting models provide systematic descriptions of international conflict behavior, at the individual, dyadic, and system levels of analysis. Two major sets of conclusions emerge. First, we need to exploit information from multiple levels of analysis to understand the initiation of any individual conflict as well as to be able to understand the general level of violence throughout the international system. Second, not all theoretical arguments are equally useful or valuable. While this may seem an obvious point, it has failed to carry the day in the field of international relations theory. Some arguments are simply more powerful than others, particularly in terms of being able to predict the relative likelihood of future wars. For instance, recalling figure 3.3, the three most powerful factors associated with the onset of war are not among the five most commonly cited arguments in the literature. Many arguments such as balance of power assertions that are widely accepted as fact provide little if any systematic purchase on predicting future risk of violent conflict within the international system. By sorting out fact from fiction, we can also tease out limited, but carefully considered, policy recommendations based on the results of our models that summarize the nature of interstate conflict over the past 175 years.
The Need for Multiple Perspectives and Levels of Analysis

We found that a large number of factors suggested by multiple perspectives appear to influence the relative risk of international conflict, thereby simultaneously giving support to many conjectures of international conflict. Rather than there being any single dominant model of international politics, as advocates for some paradigms or perspectives maintain, multiple explanatory factors from multiple levels of analysis are necessary to understand best international conflict. Contrary to some claims, the wars we observe are not overdetermined—no single factor or explanation is fully adequate to understand the onset of large-scale war. None of the individual arguments we examined comes close to explaining the majority of international conflict or even a substantial proportion of it. Moreover, there does not yet appear to be any single grand theory that incorporates the predictions of the others. Instead, the key variables suggested by most approaches have relatively comparable and relatively weak effects on interstate conflict. This is an especially important observation when considering expected utility theories of war, sometimes presented as a complete data-generating process—an all-encompassing approach to the study of international politics. While it is theoretically possible to fit arguments about regime type, the balance of capabilities, economic expansion, and so forth into an expected utility framework, to understand the initiation of disputes and wars requires a wide-ranging combination of factors drawn from multiple analytic levels, something no extant expected utility model accomplishes. Recent work in bargaining theory holds out great promise in this regard (Powell 2002; Reiter 2003).

From the perspective that we need to understand the heterogeneity among our cases, multiple interactions of casual factors, and cultural context, a long bill to be sure, someone skeptical of statistical work might question whether an alternative to our approach, namely, a qualitative analysis, holds the answer. Unfortunately, while the careful analysis of a small number of wars, such as those studies focusing on the origins and prosecution of World War I, provides better opportunities to identify the causal processes that lead to the war in question, it does not allow us to consider baseline conflict rates or increases in the relative risk of war as we have done here (King, Keohane, and Verba 1999). Political scientists working in this descriptive tradition tend to see events as overdetermined compared to historians, who tend to take a more nuanced view of causality and motive (e.g., Van Evera 1999). Nor does a qualitative approach allow us to estimate the relative effects

203
on the risk of war associated with variables drawn from the population of possible conflicts while simultaneously controlling for the effects of competing explanations.

The Relative Power of Theoretical Conjectures

While we find that multiple perspectives are necessary to explain any single conflict, we have also demonstrated that we need not weigh the empirical contribution of all these arguments equally, as the scale of the associated effects on the risk of war ranges widely. While none of the factors has what we would consider a dominating effect associated with it, the typical “real world” change in the risk of conflict associated with our different indicators is much larger for some than for others. All the same, in the context of a baseline risk of war of 1 per 14,000 dyad-years, even the most powerful individual explanation, systemic power concentration, raises the associated risk of war to a level no greater than about 1 per 1,000 dyad-years. Only when the effects associated with multiple perspectives accrue does the estimated risk of war mount to levels where we intuitively sense a high level of risk or danger. For the most part, statistically based explanations of international politics do not focus much attention on this type of interaction. Given the weakness of the relations between the operational indicators and the onset of war, our results suggest that increased attention to how multiple incentives and factors combine to increase the risk of conflict may prove fruitful. Based on our analysis, this path holds more promise for predictive power than a focus on any individual argument or conjecture.

Our analysis demonstrates that we have sufficient cases and data to estimate carefully the relative empirical effects of multiple variables on the outbreak and escalation of international conflict. However, we have also gone beyond simply recognizing that multiple variables contribute to international conflict to presenting a methodology and a set of findings on the relative effects of those variables on conflict. The directed dyad-year framework provides the ability to extend these analyses in the future to include a host of other factors from the various levels of analysis. Here, our models allow us to incorporate information from multiple levels of analysis to predict either dyadic conflict behavior or aggregate conflict throughout the international system.

We find that both system power concentration and the democratic peace proposition remain among the most powerful predictors of war and peace, even after controlling for a greater number of alternative explanations drawn from multiple levels of analysis than those found in
previous tests. This suggests that previous results are not due to model misspecification or the examination of a limited region or period as claimed by some (Gowa 1999; Henderson 2002). At the same time, however, we find that variables used to operationalize the expected utility equilibria are not among the stronger empirical predictors of the likelihood of conflict as claimed by proponents of the theory (Bueno de Mesquita and Lalman 1992). In our models of war onset, dyadic democracy is the most powerful dyad-level explanation we find.

Policy Implication for the Epistemological Optimist

These types of predictions may be of some use in the field of policy analysis, where cost-benefit frameworks are becoming increasingly important. It is true that many of the factors identified, particularly those that have the most powerful influence on the conflict behavior of states, are either immutable (contiguity), shift at almost glacial pace (system power concentration), or appear to present something of a Faustian bargain (global economic upswings in development and wealth associated with increased risk of increasingly violent war). Nevertheless, we find solace in the fact that there are a few results with direct policy implications. The first and most obvious is the democratic peace finding, even though, as always, a caveat is in order here.

Since the mechanism that appears to lead to the absence of large-scale war between democracies remains elusive (whether it exists in their electoral constraints, in their shared norms, or in their similarity of interests), we need to be prudent about advocating plans to encourage the spread of democracy on security grounds. Note that we do not suggest prudence because we find the process of democratization to be destabilizing. Like Ward and Gleditsch (1998), our results strongly refute Mansfield and Snyder’s (1995) argument that “democratization” per se is dangerous. Rather, recognizing the inherent pitfalls in the democratization process (Przeworski et al. 1999) and that the resources to support this type of enterprise are limited, the leading democratic states should place their democratization or state building bets carefully. The best gambles, and they are gambles to be sure, would be not only in those situations where the democratization efforts are likely to succeed but also in those areas suffering from the greatest risk of costly war. From the democratization literature, we know that for these pledges to succeed they must be made with long-term commitments to the development of stable democratic institutions and liberal political and legal cultures.
Our findings on system power concentration might be off-putting to some, as it would appear to be a gloomy result in these days of exceptional American economic and military power (Brooks and Wohlforth 2002). As Keohane (1984) notes, however, international institutions can play a powerful role in tempering the incentives for the use of force. These institutions may take the form of dyadic or multilateral defense pacts as we tested here, or they may take the form of multilateral economic trade institutions, whose indirect effects we also tested. While high levels of systemic power concentration create a relatively riskier environment, international institutions that attenuate incentives for the use of force can more than offset the higher levels of risk attributed to system-level characteristics. Institutions that restrain arms races, for example, belong in the category of policies worth advocating on both normative and positivist grounds.

When the United States moves to pull out unilaterally of long-standing arms control treaties, policy choices that threaten to reignite long-dormant arms races, the evidence clearly suggests that this raises the overall risk of conflict and war rather than reducing it. We also note that containment policies designed to create some sort of equitable regional balance of power in areas where conflicts loom are likely increasing rather than decreasing the future risk of war and violence. Rather than creating situations where all sides feel they have a reasonable chance to win a war should one occur, a more sensible policy would be for a powerful state or institution such as the United States or NATO to pursue a policy of extended immediate deterrence, along with concomitant building of democratic institutions. Consistent with work elsewhere (e.g., Huth 1988), potential aggressors seeking to alter a regional territorial or policy status quo are much less likely to escalate or initiate war if they fear losing that war. Creating locally equal balances of power increases these risks, while making credible deterrent commitments and establishing a clear leader reduces them.

For the scientific optimist, the bottom line is this: we need multiple theoretical perspectives to understand the onset of war, but not all theories or arguments are as valuable as others. Those whose scientific commitment is to comparative hypothesis testing as the path to new knowledge can take solace in the fact that progress is, in fact, being made. Moreover, our ability to anticipate future events improves by explicitly incorporating information from multiple levels of analysis simultaneously. However, this sort of monitoring of the scientific tote board is quite different than reaching a profound understanding of the causes of war, the principal goal of both Richardson and Wright. From this perspective, our analysis provides reasonable ground for serious skepticism.
A MORE SKEPTICAL INTERPRETATION

In our minds, there are two possible viewpoints from which one could judge the contributions to the state of knowledge made by this work. One view, that of the classical hypothesis tester, is quite optimistic. As noted previously, we have made great strides in winnowing through a vast sample of the potential correlates of war. That said, most of the conjectures we tested have no direct linkage to the causal processes generating the correlations we measured, a point that leads to our second perspective. The main goal from this viewpoint is to understand the underlying causal mechanisms that make war a more or less likely proposition. Without a carefully stated theory of why the factors we tested are associated with varying risks of war, we cannot be sure that we are not spuriously attributing causality where none really exists. An observer from this second viewpoint might have reason to be quite pessimistic, in part reflecting the skeptic’s view that the outbreak of war is a seemingly random event or suggesting (as did Richardson) that chance powerfully tempers our knowledge about the onset of any particular war. In this view, our abstract models of organized violence amount to little in the way of systematic knowledge about the nature and timing of future violence. The person interested less in narrow hypothesis testing than in identifying the causal processes that lead to war may likely come away from this work dubious about the nature of our findings. This person is likely to conclude that we still know very little about the likely origins of individual wars and could reasonably be pessimistic about our ability to generalize about the causes of events, either those that have run their course or those that may occur in the future.

Running in Place?

The skeptic would note that in many ways we are still where we were twenty years ago in the study of causality and international conflict and that cumulative progress has been numbingly slow. This book points to three empirical conjectures associated with exceptionally strong explanatory potential: power concentration in the international system, arms races, and the democratic peace. In each of these cases, we have made little theoretical progress in the past twenty years toward an accepted understanding of the underlying causal processes involved. For example, the basic arguments found in the modern arms race literature draw directly from Huntington’s work in the 1950s. The argument that the international system structure provides powerful incentives for or
against the use of force is perhaps the oldest proposition in the study of world politics. The most widely accepted finding here, the democratic peace proposition, dates to the 1970s. Although there has been a growing volume of work investigating the underpinnings of the association to identify the causal processes at work, there is as yet no consensus on why liberal democracies have not gone to war with one another.

The Limits on Testing Explanations

While one can find solace in the fact that we succeeded in identifying numerous empirical regularities—that is, sets of variables associated with increased and decreased risks of conflict—if one is interested in gaining a deeper understanding about the theoretical underpinnings of the international system, then one would rightly be skeptical about the findings and implications of our analysis. The skeptic might note that, rather than presenting a grand test of myriad theories against one another, we have simply succeeded in testing a variety of empirical propositions linking operational indicators to the incidence of interstate conflict rather than testing the causal paths of even a single theory. From this perspective, then, the findings here constitute at best a road map of facts about the nature of conflict in the international system. And while we would argue that having such a road map is critical to making scientific progress, the skeptic would note that this analysis does not directly lead us toward a satisfactory explanation of the competing causal linkages implicit in the arguments, models, and conjectures we examined.

For instance, our empirical measures do nothing to help us evaluate the competing notions of liberalism found in the various electoral constraint mechanisms and the elite-level, norm-based explanations of the democratic peace proposition. In another example, regarding the association between military spending and the risk of war, while we find that arms races approximately double the risk of war, we have no means to distinguish between the psychological stories of arms races leading to security dilemmas (Jervis 1976) and various rational choice explanations of the association (Morrow 1989; Kydd 1997). Similarly, while we find a reduced rate of interstate conflict during the cold war as predicted by Waltz (1979), before jumping on the structural realist’s bandwagon we note the inferential problem that the collinear and competing explanations of NATO institutions and nuclear weapons provide. Our inability to sort out these competing explanations must temper any claims about the actual predictive power of Waltz’s parsimonious model. The
committed skeptic could easily raise similar questions about any of the relationships identified in our empirical models.

Unfortunately, from the skeptic's perspective, sorting out where to go next is not a trivial task. When we cannot identify or measure causal factors directly, the observable implications of an argument about the origins of war may be quite distant from the supposed core arguments of our “theories” (Fearon 1996). If some variable does not fit the data in an analysis as expected according to some theoretical argument, it could be because of a flaw in the underlying causal chain, because of errors or omissions in the logic of the argument, or because of measurement error associated with some key concept. As we suggested earlier, we believe that such a tack must begin with greater care in theory development. For the most part, what international relations theorists call “theories” should be more accurately thought of as descriptions of a relatively few number of cases, and as such most of what constitutes international relations theory deserves the more precise label of conjecture, inductive empirical proposition, or simply “hunch.”

From the optimistic perspective, we might think that this type of work is similar to medical research on the things that affect human health. There, the risk of disease across the population is commonly low, as is the case here with the risk of war. The beneficial effects associated with various treatment strategies are initially unknown and probabilistic in nature. Medical researchers also have a poor theoretical understanding of why many drugs work in some situations and not in others. Yet hope springs eternal in the field of medical research. Why should we not share the same sort of optimism about future payoffs with this type of large-\(n\) research? The key difference between research searching for new medical treatments and the efforts here is our inability to run experiments. The notion of random assignment of treatment—where some patients receive a drug and others receive a placebo but neither the patient nor the physician knows who received what—is the essential tool that separates the faith we have in so much of medical research and the skepticism we hold for so much of social science research. The ability to run randomized controlled experiments means that medical researchers do not need a real understanding of the causal mechanism behind improved human health in order to be able to judge the efficacy of a given drug or treatment. International relations scholars, limited to simulations and quasi-experiments, do not have the luxury of proceeding without a clear understanding of the causal process that generates the cases we study.

Unfortunately, because we are unable to run experiments on nation states, the only way we have to test conjectures about the onset of war
or disputes is to look for real-world relationships between our dependent variables and the operational measures used as proxies for the various causal factors hypothesized to influence the onset of war (King, Keohane, and Verba 1999). With the statistical tools employed here, while we cannot directly examine many of the steps in some argument’s alleged causal chain, the correlations between our final proxy variables and outcomes are all we have to go on. For the most part, the alternatives are worse. Counter to Walt’s (1999) recent claims otherwise, even if some theoretical argument is useful as a heuristic tool for understanding previous and otherwise apparently idiosyncratic events, if its predictions do not fit to some systematically drawn random sample of the data, we must question its value for the purposes of creating new knowledge and making scientific progress. Aesthetics hold little value for us when it comes to judging the value of competing explanations of world politics.

Directly related to the problem of not resolving competing theoretical arguments is the problem of evaluating empirical assertions for which there is simply not enough evidence to judge which claim is correct. For instance, it is difficult to sort out the effects associated with NATO, nuclear weapons, and bipolarity on disputes that escalated to war during the cold war. With no wars occurring between either NATO members or the nuclear states, which could only have occurred during the post-1945 era, we cannot obtain reliable estimates of the effects associated with these factors. There is simply not enough evidence in our data set to judge what the source of peace is between the nuclear and/or NATO states. This is true regardless of the quantitative methods used to test the propositions, as there is not enough variation in the relevant variables for any technique to be able to judge the separate and independent effects of these factors. In these instances, we need either new indicators that will yield sufficient covariation to test the conjectures separately or new theories and appropriate data cast at a finer level of analysis. While we have established that there is a problem judging among these explanations, the skeptic notes that, given the tools at hand here, we cannot provide a solution.

While we find support for some arguments, the skeptic also notes that some key variables drawn from the most widely cited arguments do not have associated with them the empirical effects that many might otherwise expect. We recognize that in a probabilistic world no single test can provide compelling evidence for or against a casual argument. However, we believe that it is important for advocates of models that failed in our tests to explain why the empirical measures for their models have little apparent association with the risk of international conflict.
after we control for other explanations. Most notably, this is the case with our measures of power transition logic, for example, and the classic realpolitik argument that a bilateral or dyadic balance of power will restrain conflict. Both of these widely cited arguments receive no empirical support in our analysis. These findings beg an explanation from traditional realists because they undermine a basic assumption driving a large portion of the security studies literature and policy. It is unclear, for example, how debates about “balance of power systems” or about whether states “balance” or “bandwagon” can be resolved when the most basic premise of a dyadic balance of power (that an equal dyadic balance restrains conflict relative to imbalance) receives no support. The divergence between empirical regularities or lack thereof, arguments that refuse to go the way of the white whale, and our minimal understanding of the underlying causal processes leads the skeptic to question how much cumulative progress international relations “theorists” have made over the past forty years. The optimist notes that we now have a better established set of facts for theorists in various traditions to address. To the skeptic, though, the point that we are still establishing such facts rather than divining what the underlying causal processes are bodes ill for future progress.

Certainly, work remains to improve the econometric model developed here; it is far from perfect. A skeptic would suggest that the statistical models’ inaccurate and relatively weak predictions are to be expected. Indeed, it may be impossible to improve them greatly, even if we dig more deeply into the data using alternative statistical or computational techniques. In one extension to our analysis, and in response to Smith’s (1999) argument that strategic selection bias might plague our results, we replicated our basic analysis using a strategic Bayesian estimator of his design. After dedicating over sixty days of continuous computer processing time to running this large model, we generated new results that differed only in small degree from those presented here.¹

A skeptic might also find it ironic that, while we have focused our attention in this book on estimating a more complete econometric model of interstate conflict, we end by suggesting the need for more theoretical work. Improved theory may help to enhance the statistical model by suggesting nonobvious interactions and nonlinear specifications of the existing variables. Similarly, we need more careful theorizing about the heterogeneity among cases, a serious and related confounding problem about which we should be much concerned. While the many structural-level explanations of international politics assume that “states are states” and that the effects of systemic anarchy swamp the causal effects of the varying internal characteristics of
states, we hope the findings from chapter 6 will disabuse careful readers of this view.

Aside from the problem of unit heterogeneity, there are several possible reasons why, in the end, our statistical models suffer from serious limitations in predicting the onset of violent interstate conflict. Some of these reasons are typical of those caveats accompanying empirical analysis in any field. For example, we simply may not have the right variables in our model. Scholars may not have found the right measures or functional forms to specify accurately some particular expected relationship. Improved measures or instruments might improve the fit of our model to the data significantly. Another possibility is that we have not yet adopted the right theoretical perspective in our analyses and that some other explanations would produce the very strong findings about conflict we hope for, findings that would dramatically improve our predictive accuracy. Finally, our proxy variables for deeper concepts may be inadequate and fraught with error. Random noise in our measures can only hurt our model’s fit, attenuating our parameter estimates.

None of these reasons is particularly satisfying, however; they sound to the skeptic’s ear more like rationalizations rather than directions for future research. In our models, we included a major cross-section of the rigorously testable conjectures about the sources of interstate war. Most of the measures we used have been subject to multiple rounds of peer review and improvement for over twenty years. It is certainly possible that there is some dark-horse explanation or measure lurking somewhere in the literature, but this appears unlikely. Of course, it could be that there is not a single omitted strong argument but that several other theories or conjectures yet identified would each explain a little bit more of the conflict puzzle.

POSSIBLE WAYS OUT OF THE TRAP?

While the work in the preceding chapters continues the empirical tradition of the past sixty years, it also suffers from its inherent limitations. Given over a million observations and dozens of independent variables, might we reasonably have expected stronger results and predictions? Two powerful alternative tools hold much promise for the type of work executed here. On the empirical front, neural network models are quite promising because they offer the opportunity to estimate unconstrained models that allow for large numbers of interactions and nonlinear relations between the independent variables and the dependent variable. On the theory front, game theory offers a rigorous
and deductive process to build real theory, something sorely lacking in the field of international politics. These complementary approaches are not without their flaws, however.

Neural Networks

One powerful empirical approach recently deployed is to use neural network models to find highly interactive sets of variables that together appear to predict outcomes relatively well (Beck, King, and Zeng 2000). Even this highly complex estimation method “is not a panacea” (22); while the method is quite effective at identifying patterns in the data, it allows us to explore the theoretical or casual structure of the conjectures included in the model only with great difficulty. We wonder whether using a largely unconstrained model is in fact any better for producing an understanding (rather than simply a complex description) of interstate conflict than the conjectures and hunches of which we have been so critical to this point. Based on this research, the best way to improve our understanding of the origins of war may not lie in the development of ever faster computers or more efficient and advanced statistical estimators. Rather, future advancement will most likely result from more detailed archival analysis of the data-generating processes that lead to the wars we observe, as well as those that never occur.

In addition to these basic problems, there are other, more fundamental reasons why this type of large quantitative analysis yields explanatory and predictive power that is lower than we would like. In presenting the advantages of neural network models, Beck, King, and Zeng (2000) argue that explanations for international conflict are likely to be highly contingent and interactive, demanding particularly close attention to interactions of circumstances and casual factors. If this is the case, then the approach of including multiple independent factors in additive fashion will surely prove inadequate. Unfortunately, systematic pursuit of the neural network approach confronts a fundamental limit to knowledge, the problem of limited degrees of freedom. Depending on how one defines the event of “war initiation,” there have been roughly seventy-five to one hundred of them in the international system over the past 175 years, yielding a terribly skewed distribution with a small number of events to explain relative to the large number of nonevents.

When we begin to increase the potential interactions between our variables, the relatively small number of wars necessarily begins to limit the number of explanations we can consider simultaneously. In their
neural network model, Beck, King, and Zeng (2000) included measures for alliances, geographic contiguity, dyadic balance of power, democracy, and the number of years since a previous conflict. Their model predicts well, in some ways better than the one presented here. For computational reasons, however, if they were to include measures for all the factors we know are systematically associated with war, they would quickly run short of the degrees of freedom needed to solve their model’s simultaneous equations. This is because the neural net approach voraciously consumes data while sifting through a myriad combination of nonlinear interactions among the independent variables. Unfortunately, while the predictive performance of their model is relatively high, we know from our results that our ability to infer about the relative explanatory power of any one conjecture is highly contingent on what else is in, or out of, the statistical model. Similarly, if we believe that there is systematic variation in the fit of each explanation across regions and time periods (as our results in chapter 6 strongly suggest), we would need to include additional interaction terms. However, even if we could mark each argument with a single variable (something we know is not the case), we would quickly end up with significantly more variables than wars (5 regions × 3 periods × 16 conjectures = 240 variables).

In part, because most of our so-called theories are in fact inductive descriptions of a small number of events, we may never possess enough data to sort out empirically which set of variables and interactions among them provides the “best” fit to the data in general. As we noted earlier, knowing that democracy correlates strongly with the incidence of conflict is far different from understanding why this is so. Models exploiting large numbers of complex interactive measures (such as neural network models) help tremendously in establishing more fully the sorts of facts we lay out here. That said, these large, brute force attacks on the problem are not as helpful when it comes to answering theoretical questions, such as sorting out why democracies have different conflict propensities than other types of states or why NATO members have never gone to war but have engaged in mutually violent disputes just short of war. They can help us find and confirm empirical patterns, but they are not as well suited to helping us understand the theoretical mechanisms underlying these relations.

Is Game Theory the Solution?

Might the solution to these many problems, the majority of which are rooted in poorly or loosely specified theories, lie in more widespread use
and development of formal mathematical models? Perhaps, but scholars building formal deductive models have not yet solved the puzzles found along the way in the quest to identify the true causes of war. Rational choice models may be subject to inherent epistemological limitations similar to those from which the brute force empirical approach suffers. A final limit to knowledge—a speed limit in a way, in terms of what we may be able to predict, particularly if the actors behave in perfectly rational ways—derives from Riker’s (1980) pessimistic quip noted at the beginning of this book. Riker was pessimistic about the future ability of political scientists to predict successfully the behavior of strategic actors because he believed there might be no fundamental equilibria to identify. Following the logic of Arrow’s theorem, Riker deduced that if there are multiple actors, with differing preferences, then there is no single equilibria, or core, that we can identify as clearly superior to others. How does this matter? In a recent paper, Lewis and Schultz (2002) show exactly how. They begin by building what we might think of as the world’s simplest strategic signaling game—a game in which the two actors predicate their decisions to use force against the other on the “signals” one actor receives from the other actor about the other’s intentions and expectations. From their Monte Carlo simulations, they reach the following somewhat unhappy conclusions:

First, regardless of the solution concept employed, non-innocuous identifying restrictions must be made. For example, the effect of a covariate on a particular payoff generally cannot be estimated unless that covariate is assumed to have no effect on other payoffs. Second, the distribution of payoffs implied by a given set of outcome data strongly depends on the solution concept employed. Thus, even in very simple settings making inferences about payoffs from data on the outcomes requires strong and untestable identifying assumptions. This problem is exacerbated if the information structure is not known, in which case assuming the wrong information structure can lead to entirely misleading conclusions about payoffs and the effects of covariates on those payoffs. (Lewis and Schultz 2002, 1)

The issue is that different game theoretic solution concepts (Nash equilibria, perfect Bayes equilibria, quantal response equilibria, and so forth) can all lead to different equilibrium outcomes. Worse yet, these different equilibria are sometimes associated with wildly different comparative statics relationships among the variables that affect the likelihood of which equilibria will emerge, given the same game tree and
similar sets of preferences. A solution concept is like a set of rules we need to follow to solve the game. The problem is that each of the solution concepts they explore is plausible—there is no compelling reason why we should prefer the Nash equilibrium concept to McKelvey's quantal response solution concept, for example. The real-world analogy is the following. Even if the two sides in a dispute are rational and agree that they face the same known sequence of choices, if they make their decisions in only slightly different but equally rational ways—in effect employing different equilibria solution concepts—then they may reach quite different conclusions about what is their best strategy. By “best strategy,” we mean the best response to what they will rationally deduce the other side in the dispute to do. As a result, it becomes difficult both for the decision makers and for us as analysts to know what solutions present the best approximation of how decision makers may act in some particular situation. Empirically, we could test which sets of equilibria match real-world behavior best to help us understand which “style” of solution concept leaders tend to use most frequently. Somewhat ironically, however, that would be employing an inductive solution to a deductive trap—an approach of which advocates of game theory in particular and formal theory in general have been highly critical (e.g., Morrow 1993; Morton 1999). This is an area where experimental methods may prove tremendously helpful.

From a skeptical perspective on trends in the discipline, it appears that the promised gains from game theory and formal mathematical approaches remain as elusive as ever, assuming our interest is the generation of testable empirical propositions that illuminate novel and powerful causal arguments. While tremendous gains in computing power and software developments such as EUGene allow us to test arguments on the entire population of dyad-years (over a million observations), the payoffs from ever more sophisticated and complex models such as Bueno de Mesquita and Lalman's (1992) IIG are not clear. If anything, the most recent sophisticated work bodes ill for future progress in terms of game theory's contribution to our improved understanding of interstate violence (Lewis and Schultz 2002).

In the end, there may be actual limits to the degree of knowledge that we can distill from empirical observation of the events we wish to study. Following Gartzke's (1999) argument as developed in chapter 2, it will likely prove impossible to predict precisely the timing of individual international conflicts. If the incidence of war is a function of unobservable private information, even in part, then as analysts we can predict, at best, some relative likelihood of observing the initiation of war. If the theoretical upper limit of this forecasting probability is fifty-
fifty, or the flip of a fair coin, then the pseudo-$R^2$ summary statistics produced with our two models of 0.26 and 0.20 are quite good. This would likely be the case if state leaders were rational, if private information was the sole cause of conflict, if outcomes were win/lose, and if the choice confronting them was war/no war. These measures of aggregate model performance would then indicate that we are predicting roughly half of all that we can know ex ante about conflict initiation (again, assuming strategic and rational actors). If the decision makers are not rational, their behavior might be completely random, further lowering our chances of successful prediction. Alternatively, they might be irrational but in completely systematic ways, thereby raising the upper limit of predictability. While the view from the perspective of strategic rationality is not optimistic about anyone’s ability to predict empirical behavior accurately, it may provide an understanding of the limited successes seen here.

WHERE TO GO NEXT?

Our results further point to focusing on differences between and interactions among states, their local and regional environments, and domestic politics as areas that need theoretical development. While recent work on the relationship between states’ varying domestic political institutions and conflict behavior begins to relax the assumption of the state as a unitary actor, the inclusion of a single variable marking state type does not come close to capturing all of the potentially important differences across states and regions. Our results in chapter 6 demonstrate tremendous variation in the observed association between the expected utility measures and conflict behavior across regions and time periods. In results not reported here, we find similar patterns for variables drawn from balance of power approaches. Green, Kim, and Yoon (2001) have argued that the effects we associate with dyadic democracy, one of the more powerful pacifying factors, are also subject to substantial spatial and temporal heterogeneity. Related to our findings on trade and war, in another recent example, McGillivray (1997) shows that, if we want to understand tariff levels and states’ general willingness to cooperate, it is not enough to account for regime type coded as democracy-autocracy. Nor is it sufficient only to distinguish between different democratic systems such as parliamentary or presidential systems (Garrett 1999; Busch and Reinhardt 2000). To understand the general demand and supply of protectionism—something going far beyond a narrow conception of a single state’s foreign policy—we also
need to understand the geographic distribution of industrial assets and actors in combination or interaction with a sophisticated notion of regime type. The weak and varying results we present for the expected utility model suggest that McGillivray and others who have developed sophisticated theories about the nature of government and private institutions and how these institutions interact with the geographic distribution of business and citizens’ interest groups are clearly on the right track (Brooks 2001).

To those working in the field of comparative politics, it likely comes as no surprise that those who study international politics need a more sophisticated understanding of regime characteristics, regional differences, and the interaction of general factors with a more clearly specified notion of local contextual factors, otherwise known as “culture.” Unfortunately, many international relations theorists choose to ignore the ways that the domestic affairs of state may alter or affect the policy choices we observe, using the presence of systemic anarchy as a basis for their claim that they can treat states as black boxes. Staking out an extreme view in this regard, Waltz (1979) and his followers argue that any theory of international politics that delves into the nature of domestic political systems immediately devolves into a theory of foreign policy and not international politics. If anarchy is largely what states make of it, however, this assertion is inadequate—moreover, recent work demonstrating the systematic effects of domestic politics on levels of conflict across the international system suggests that this assertion is patently wrong.

These pessimistic limits to our ability to forecast the future notwithstanding, with the analysis here, we provide a benchmark for further comparative tests of the power of multiple explanations for international conflict. Analysts who believe that different measures or additional explanations of conflict need to be included in analysis can easily extend our work using the software and replication data available at www.eugenesoftware.org, replacing or adding measures to reflect differing theoretical concerns. The key here is that the analysis of other variables and their permutations, interactions, and explanations proceed not in isolation but side by side with the multiple factors already known to influence conflict behavior.

We see at least three immediate directions for future research to expand upon our findings. New inquiries might fruitfully progress by seeking systematically to develop theoretical understandings of the heterogeneity among different dyads or regions we demonstrated in chapter 6. At this point, we do not have a good understanding of how we should conceive of international heterogeneity, even at the most basic
level. States are all members of the international system, but they are simultaneously members of other types of systems as well: systems of regions, alliances, and dyads. As we showed in chapter 6, rational choice models of international politics fit the data quite differently across regions, time periods, and length of interaction; the same is true for other theoretical approaches such as liberal institutionalism (Cederman 2000, 2001). However, the appropriate level at which to model heterogeneity is unclear. If every dyad has a distinct baseline propensity for conflict, then we would want to use fixed-effects models for our estimations (Green, Kim, and Yoon 2001).

Related to this point are issues of selection bias in our studies of international politics. For example, analyzing samples known to be heterogeneous (such as “politically relevant” dyads), without understanding the selection mechanisms that generate such samples, risks producing parameter estimates that we cannot confidently use to make predictions across states and time. We need to consider whether and how our conflict models should apply to all cases or to nonrandom samples identified before the fact and how to integrate theoretical arguments that only apply to selected cases (such as power transition models of great power war) into the general approach we have advocated here.

Another important direction is to develop further a systematic understanding of how different models and explanations of international conflict interact. The infrequent confluence of a multitude of otherwise weak predictors of war may serve to explain the relative rarity of international conflict, as the large increases in the predicted risk we observed when combining changes in multiple variables suggest. It may also be the case that an additive approach such as the one we used here is simply inadequate to capture the contingent nature of most causal relationships, and we should therefore include a wider variety of interactions in our statistical models. For example, although we employed contiguity as an additive control variable, we may instead wish to include it as an interactive factor marking a concept such as “opportunity.” However, we believe it is more appropriate to approach this issue from a theoretical perspective first rather than simply exploring the data or running saturated models to find variable interactions that are empirically powerful at predicting conflict in our data.

A final issue concerns the inclusion of more dynamic factors in our analysis. Our model does better at identifying conflict-prone dyads than it does at identifying the point in time when the members of a particular dyad will turn to war. One possible reason is that the models and hypotheses we tested are largely static and that the indicators included...
are largely structural rather than dynamic. Structural indicators are those that do not change as a conflict develops, while dynamic ones are those that can change rapidly, more accurately reflecting the policy decisions and choices that lead to the disputes and wars making up our dependent variable. Structural indicators are effective in predicting the behavior of a system if and only if the system is primarily in equilibrium (or close to it). However, if most social behavior is far from equilibrium, or has no fundamental equilibrium at all, then structural indicators will tell us relatively little about how the system operates in terms of the timing of individual events, which is the information we desire most.

Most of the variables in our statistical models do not change rapidly (e.g., capabilities), or else when they do, we often do not observe the changes measured on a time scale that will allow us to either make precise predictions about conflict timing or sort out the precise causal sequence leading to the outcomes we observe. For example, the fact that we must lag most of the variables we measure annually in order to avoid measuring factors after a conflict has already occurred contributes to this problem. Largely immutable indicators, such as polarity or power concentration, plausibly represent the structure of the system at the time of a series of dyadic interactions. As a result, they may affect the baseline risk of war, but they do not provide the day-to-day (or month-to-month) dynamic actions, reactions, and stochastic shocks to the system that represent critical flashpoints and proximate causes of violent conflict.

CONCLUDING THOUGHTS

In the end, what does all this mean for our general understanding of international politics? On the theory of international politics front, we have answered far fewer questions than those that remain outstanding. However, we have been able to present an assessment of the relative predictive power associated with a large variety of factors (some manipulable, some less so) purported to be causes or inhibitors of interstate conflict. We can also conclude that turning exclusively to a small-$n$ approach will prove inadequate as an alternative to the kind of analysis we performed here. This does not mean either that we can rely on large-$n$ studies to supply us with satisfying answers to the important questions about the theoretical bases of international politics. Instead, qualitative studies can provide a powerful counterpart or complement to this type of work; indeed, large-$n$ statistical studies in the absence of
Conclusion

controlled experimentation cannot provide the fine-grained information we need to understand fully events in the past as well as those that have yet to occur. It will likely be in detailed archival analysis that we will make progress in identifying the underlying causal mechanisms we have yet to put to any real form of test.

We close with a call for methodological pluralism. The skeptic is correct in noting that naive empiricism cannot provide the information we need to understand the causes of war. At the same time, our statistical optimist also rightly points out that purely historical research does not let us plausibly judge how the risk associated with one situation compares to the next. Moreover, we cannot rely on theory alone either; we also know that indeterminacies often riddle the equilibria derived from game theory models, the most rigorous theoretical approach we have at our disposal. The research efforts that will move us forward in the next era of peace research will likely combine aspects of all three approaches to the creation of new knowledge. Research into the causes of war has matured to the point where all of the comparatively easy tasks are done. Next comes the truly difficult work, work that will combine formal mathematical theory with archival work to establish causal pathways and complementary large-\(n\) tests where the data are available to enable carefully controlled comparisons to estimate the relative risks associated with the various indicators of interstate violence.

Looking with a retrospective view over the past sixty years of quantitative research, it seems clear that we have learned a great deal and have established several important facts relating arms races, domestic politics institutions, and the nature of the international system to the onset of war. It is also clear, however, that chance plays a powerful role in determining the individual events that shape our history and that many of the known leading indicators of international conflict have individually weak effects associated with them. As Bismarck noted in his memoirs, a ship’s captain must keep an eye to the weather, as he cannot dictate the winds but rather must react to them. There will be times when the port of choice is out of reach because of the direction of the prevailing winds, and there will be times when favorable winds will carry a ship farther than anyone might have anticipated in advance. The truly wise ship captain also recognizes that, even with the best ship and finest crew, there will be times when no port offers a safe berth and that, at those times, prudence dictates sailing into the wind and waiting out the storm.