In *Theory and Evidence in Comparative Politics and International Relations*, ed. by Richard Ned Lebow and Mark Irving Lichbach. New York: Palgrave Macmillan, 2007. Pp. 177-97.

7

Theory, Evidence, and Politics in the Evolution of International Relations Research Programs

Jack S. Levy

The field of international relations has always been diverse in its metatheoretical and methodological orientations, perhaps more so than any other field in political science, and intrafield debates about the proper way to study world politics has made it a richer, more interesting, and stronger field. The contentious nature of the discipline is reflected in the fact that the history of the field is often told in terms of a sequence of "great debates." These include debates between interwar idealists and postwar realists (Carr 1939; Morgenthau 1948), between "traditionalists" and "behavioralists" in the 1960s (Bull 1966; Kaplan 1966), and among realists, liberals, and Marxists beginning in the 1970s (Gilpin 1975; Guzzini 1998).

Until recently, these debates were conducted within certain limits (Holsti 1985). Despite their differences, most traditionalists and behavioralists adopted a realist world view (Vasquez 1983; Schmidt 2000). Similarly, the "paradigm wars," particularly between neoliberalism and neorealism, were conducted within an underlying rationalist consensus (Waever 1998; Ruggie 1998). In the last decade, however, that consensus came under sharp attack by various forms of postpositivism, including postmodernism, poststructuralism, feminism, and constructivism. This so-called third debate (Lapid 1989) is in many respects more profound than earlier ones, because underlying ontological and epistemological issues are at the core of the debate.

One theme in these ongoing debates concerns the criteria by which scholars evaluate progress in the cumulation of knowledge. In the last decade or so international relations scholars have been more explicit in grounding their conceptions of scientific progress in particular approaches in the philosophy of science. Many have used Imre Lakatos's (1970) methodology of scientific research programs (Vasquez and Elman 2003; Elman and Elman 2003), while others have criticized Lakatosian metatheory and turned instead to Popper (1957, 1962) or Laudan (1978).[1] Still, each of these metatheoretical frameworks falls within a positivistic conception of social science. This has led others, including many of the contributors to this volume (Bernstein et al., Hopf, Kratochwil, and Lebow), to adopt more critical perspectives. They deal with questions of ontology and epistemology as well as method, and they attempt to broaden the conception of science and thus of what constitutes scientific progress.

While these debates focus on the normative questions of what constitutes scientific progress and the proper criteria for evaluating progress, and thus on how research ought to evolve, my own concern in this chapter is with the more descriptive question of how scientific research programs actually evolve. That is, I am concerned more with the history of research programs than with the prescriptive methodology for evaluating them.[2]

More specifically, I ask the related questions of what factors influence the evolution of research programs and why some programs or traditions are more "successful" than others, defined in terms of their impact on and endurance in the field.[3] I give particular attention to the relationship between theory and evidence. Is the research process dominated by theory, so that research programs endure because they are characterized by theoretical elegance, deductive fertility, and wide-ranging explanatory power? Or is the research process driven by evidence, with the most successful research programs characterized by the extensive support they draw from the accumulation of empirical evidence? Alternatively, do research traditions endure because they respond to current events and/or reflect the policy agendas of the governments or perhaps competing elites?

These questions are more descriptive than normative, and they lead me to direct my primary attention to the level of research design. This stands in contrast to most of the chapters in this volume, which give more attention to questions of ontology and epistemology. Admittedly, questions of method cannot be entirely separated from more fundamental metatheoretical questions. At the same time, however, a useful prescriptive methodology for how a research program ought to develop cannot be entirely divorced from an understanding of how research programs actually develop, and this chapter on the history of research programs provides a

useful perspective for the more metatheoretically oriented essays in the rest of the volume.

I organize this chapter around a simple typology of the primary factors influencing the evolution of research programs: theory, evidence, and politics. I argue that different research programs follow different paths to success (and to failure), and that these different paths involve different sequences of theory and evidence. Some research programs are primarily theory driven, others primarily evidence driven, and still others are driven by an alternating sequence of theoretical conjectures and empirical refutations.[4] I illustrate these different paths with examples from a number of research programs in international relations and in political science more generally. I then consider the impact of current events and policy agendas on research programs. I argue that some research programs evolve independently of specific normative values or policy agendas and are driven primarily by autonomous analytical developments or by evidentiary support.

Space constraints preclude a fully systematic empirical analysis of a variety of research programs and their historical evolution. Ideally, such a study would incorporate research programs characterized by variation across a number of dimensions. They would include non-American as well as American scholarship,[5] qualitative as well as quantitative and formal research, and work that falls outside as well as inside positivistic social science. Given the goals of understanding why research programs succeed, it would also be important to include failed research programs. This is not at all possible in a short essay, but our coverage will be broad enough to demonstrate the multiple paths through which international relations research programs develop.

It is useful to acknowledge at this point that the task of assessing the relative impact of theory, evidence, and policy on the evolution of research is complicated by the fact that research programs are generally macrolevel phenomena that involve many scholars and that represent the aggregation of many individual decisions as to where to focus their scholarly efforts. Different scholars may choose to work within a given research program for different, even diametrically opposed reasons. In addition, one set of factors may influence the initiation of a research program while other factors may help to sustain or expand it. These complications make it difficult to identify a single pattern underlying a particular research program.

Theory-Driven Research Programs

Some research programs are driven primarily by theoretical considerations. The strength or quality of a theory is a function of a number of criteria

(Hempel 1966), including its degree of falsifiability and internal consistency,[6] its deductive power, its elegance and parsimony,[7] the plausibility and completeness of its hypothesized causal mechanisms, the range and number of its testable implications, and its consistency with existing laws and theories that have themselves received substantial degrees of empirical support. These are scientifically normative criteria, and are best distinguished from substantive normative values, which I treat in a separate category. Note that many of these criteria are matters of scholarly convention, as Chernoff argues in his contribution to this volume.

While empirical validation of the theory's key propositions clearly enhances its scholarly impact, a research program propelled by a powerful theory can be self-sustaining, even in the absence of a significant amount of supporting evidence. The best example comes from economics, where microeconomics is dominated by general equilibrium theory and a commitment to mathematical formalism and has little empirical content (Weintraub 1985; Backhouse, 1994). The best examples in political science are associated with the rational choice paradigm. Arrow's (1953) general impossibility theorem, Down's (1957) median voter theorem, and Olson's (1965) theory of collective action; each generated enormously influential research programs quite independently of empirical validation, although empirical work on the latter two topics subsequently reinforced those programs.[8]

In international relations, one of the best examples of a theory driven research program is the "bargaining model of war."[9] This rationalist model is based on Fearon's (1995) formalization of an idea suggested by Blainey (1973) and familiar to most economists: war is an inefficient means of settling disputes because it destroys resources that could have been shared by the contending parties. The question that needs to be answered, then, is what precludes parties with conflicting interests from reaching a negotiated settlement that avoids the mutual costs of violent conflict.

This fundamental idea generated a significant line of theoretical research that focuses on the role of "commitment problems," "private information" and incentives to misrepresent that information, and the divisibility or indivisibility of issues (Gartzke 1999; Powell 2002; Filson and Werner 2002; Wagner, 2000). The model has been applied to the study of ethnonational conflict (Fearon and Laitin 1996; Lake and Rothchild, 1998) as well as to interstate conflict. The conception of war as an information-revealing mechanism has also led to hypotheses about the termination of war (Slantchev 2003), some of which have recently been tested empirically. One measure of the influence of the bargaining model is the extent to which the concepts of private information, commitment, and issue indivisibility have become prominent in the qualitative as well

as formal literature on international conflict, quite independently of any empirical confirmation of key propositions.

The bargaining model of war emerged quite independently of any obvious normative assumptions or policy issues. It was the product of certain analytic developments in game theory, particularly the incorporation of "incomplete information" into game-theoretic models beginning in the late 1980s. The bargaining model and in fact most of the contemporary game-theoretic models of economics and political science were not possible until economists invented certain analytic techniques that permitted the analysis of games with incomplete information.[10]

We could also include broader paradigmatic approaches such as realist, liberal, and Marxist-Leninist international theories as examples of theory-driven research programs, but difficulties quickly arise. These paradigms contain multiple theories that are not necessarily based on the same set of hard-core assumptions and that consequently may contain contradictory propositions. This means that evidence falsifying one theory might validate another, always leaving some theory within the paradigm consistent with any empirical observation, and thus leaving the paradigm itself immune to falsification.[11] More important, the analytic assumptions underlying each of these paradigms are much more normatively loaded than those for the rational choice paradigm (though not necessarily more than it is for specific substantive theories within rational choice, such as deterrence theory), and it is consequently much more difficult to differentiate the influence of abstract theory from the influence of policy agendas.

Evidence-Driven Research Programs

Some research programs are driven more by evidence than by theory. The strength of evidence refers to the overall quality of the research design; its effectiveness in controlling for extraneous variables and in dealing with endogeneity problems; the validity of empirical indicators for key theoretical concepts; the quality of the data; the appropriateness of any statistical methods used in the analysis of the data, including the fit between the assumptions of the statistical model and those of the theoretical proposition to which it is applied; the replicability of the data analysis and the extent and variety of replications of the analysis; the appropriateness of case selection, given the theory to be tested, including the sensitivity to possible selection bias; and the extent to which the findings can be generalized to other spatial and temporal domains beyond the immediate data.

These criteria are fairly standard in books on research methods, the most influential of which is King, Keohane, and Verba (1994). Their

Designing Social Inquiry gives more emphasis to empirical criteria than to theoretical criteria for scientific progress,[12] and it strongly suggests that successful theories are those with the greatest levels of empirical support—points that many of the book's critics have noted (Brady and Collier 2004).

One example of an empirically driven research program is the one on territory and war (Vasquez 1993; Vasquez and Senese 2004; Huth 1996; Hensel 2000; Huth and Alee 2002). This scholarship has been propelled by the repeated demonstration that a disproportionately high number of wars involve territorial disputes and that territorial disputes are more likely to lead to war than are other kinds of disputes. Thus far, however, there is little agreement on the precise causal mechanisms leading from territoriality to militarized conflict. This is clearly the case of a strong empirical finding coming first and stimulating theoretical efforts to explain that finding and associated empirical relationships.[13]

A more influential research program that is primarily evidence driven, at least in its early stages, but that involves a complex mix of factors, is the democratic peace.[14] It is undoubtedly true, as Lawrence argues (this volume; see also Oren 1995), that many joined in the study of the democratic peace because of a normative commitment to liberal democracy and to an American foreign policy agenda of actively promoting democratic values abroad. Yet perhaps just as many engaged the debate because of their realist worldviews, a determination to demonstrate the fallacy of early empirical work, save realism from one of its most glaring empirical anomalies, and steer American foreign policy away from a misguided liberal interventionism. While conceding the impact of policy agendas (liberal and antiliberal) on the democratic peace research program, I want to emphasize the primacy of another factor, particularly in the early stages of the research program—the unprecedented level of empirical support for the dyadic-level finding that democracies rarely if ever fight each other. It was the strength of this correlation, along with the absence of any unambiguous anomalies, that generated both the intellectual curiosity and professional incentives for realists, liberals, and others to redirect their research energies toward the democratic peace.

After Doyle (1983) and later a special issue of the *Journal of Conflict Resolution* (December 1984) emphasized that democracies almost never go to war with each other, many of the scholars who initiated research on this question were skeptics who were convinced that the findings were based on flawed research designs and who were determined to introduce greater rigor into democratic peace research. This certainly applies to Singer, who coauthored the first systematic study of democracy and peace (Small and Singer 1976) and who remains a skeptic. It also applies to Weede (1984), Bremer (1992), Maoz (1992), Bueno de Mesquita and Lalman (1992), and Russett (1993), each of whom is now a strong believer in the dyadic democratic peace.[15] Many of these scholars may have wanted to believe that democracies rarely if ever fought each other, but most were skeptical of the validity of the finding and expected that it would wash out once scholars controlled for other key variables such as trade, distance, alliances, and the like. It was the near law-like character of the interdemocratic peace proposition, in a field in which relatively few empirical regularities of even modest strength had been uncovered, that energized scholars to engage in further studies in an attempt to validate or invalidate the early findings, to explore potential anomalies in more detail, to consider the possible extension of the findings to earlier temporal domains and to other international systems, and to generate and test additional theoretical implications of the democratic peace proposition. As the consensus grew that the dyadic democratic peace was real, so did the professional incentives for individuals to attempt to demonstrate that the proclaimed absence of war between democracies was the artifact of misspecified theoretical arguments and flawed research designs.

Some will disagree with my emphasis on the primacy of evidence in the evolution of the democratic peace research program, and a more thorough and systematic analysis is necessary to resolve the debate. One thing that everyone agrees has had little impact on the study of the democratic peace, at least until recently, is a strong theory. The empirical finding clearly came first, followed by attempts to validate it and to explore possible anomalies, and finally by theoretical conjectures to explain it, none of which has generated overwhelming support. The relative absence of war between democracies remains a strong empirical regularity in search of a theory to explain it.[16]

The Dialectic of Theory and Evidence

The theoretical and empirical dimensions of a research program are not analytically distinct, of course. One cannot analyze the evidentiary support for a theory apart from its fundamental assumptions and propositions or from alternative explanations for evidence consistent with the theory, all of which affect case selection, the operationalization of key variables, and all other aspects of research design (Merton 1974). A research program driven entirely by evidence, without any prior theoretical assumptions, is inconceivable. Few contemporary scholars would embrace the epistemology underlying Sgt. Joe Friday's (of "Dragnet" fame) request for "Just the facts, ma'am, just the facts." As Goethe wrote, "Every fact is already a theory" (cited in Waltz 1997: 913).

To say that all research is guided by theory does not imply that theory necessarily plays a greater role than does evidence in research programs, any more than the fact that most theories are influenced by some prior empirical observations implies that evidence plays a greater role. It is not clear, however, exactly how we should assign weights to theory and evidence. The problem is compounded by the multiple ways in which scholars use the term theory—to refer to everything from axiomatic deductive theory to broader conceptual frameworks or paradigms with contested and conflicting theoretical assumptions and only vaguely specified causal mechanisms.

One example is the ongoing debate over whether a preponderance of power or a parity of power is more likely to lead to war. This dyadic-level debate grew out of the power parity hypothesis of balance of power theory and the power preponderance hypothesis of power transition theory (Organski 1968). Neither theory carefully specified the causal mechanisms leading from structure to outcome,[17] but the debate was dominated by a series of empirical studies beginning in the 1970s.[18] There is now strong empirical evidence in support of the power preponderance hypothesis (Kugler and Lemke 1996), but the precise causal mechanisms remain poorly developed, in part because power transition theory has yet to incorporate a theory of bargaining (DiCicco and Levy 1999). I see this pattern as reflecting the dominance of evidence over theory in the evolution of power parity/power preponderance debate, though theory probably plays a greater role in the debate between balance of power theory and power transition theory.[19]

It is also possible that a research program can combine theoretical and empirical elements in an alternating sequence of theory and evidence: a reasonably well-specified theory leads to empirical tests that contradict some of the testable implications of the theory, which then leads to the modification of the theory or perhaps to its replacement by an alternative theory. Or the process may begin with robust empirical findings that lead to the construction of a theory to explain them, which leads to new predictions that guide subsequent empirical research. An alternating sequence of theory and evidence fits Popper's (1962) model of conjectures and refutations.

We start with a hypothesis, whether derived from a theory or induced from observation, test it against the evidence, and use the evidence to refine, revise, or reject the theory. This idea is explicit in the methodology of structured, focused comparison (George and Bennett 2005) and in the methodology of the analytic narrative research program (Bates et al. 1998).

This strategy for the cumulation of knowledge is also influential in historiography. Carr (1964: 20–21, 26–30) criticized both Rankean historiography (Iggers 1984) for its "fetishism of facts" and historical idealism for its argument that empirical observations are entirely determined by theoretical preconceptions. Carr argued that "the historian is neither the humble slave nor the tyrannical master of his facts," and that history is "a continuous process of interaction between the historian and his facts, an unending dialogue between the present and the past." Similarly, many of the essays in this volume explicitly or implicitly accept as a normative ideal the model of an unending dialogue between theory and evidence, recognizing that theories with different ontological and epistemological foundations call for different kinds of evidence.

Still, there is a less-than-perfect fit between the conjectures and refutations ideal and the reality of political science research programs. In contrast to physics, which in many respects provides the paradigmatic case for Popper's model,[20] the social sciences provide fewer clear-cut rejections of a given theory. A possible exception are the experimental social sciences, where highly controlled experiments generate greater consensus on the refutation of theoretical conjectures.

A good example here is decision theory. If we define this research program broadly to include both formal (normative) decision theory and more descriptive research in social psychology and behavioral economics on how people actually make choices under conditions of risk,[21] then we can interpret a long line of work on decision theory in terms of an alternating sequence of conjectures and refutations.

We can trace the initial conjecture of decision theory to Pascal's proposal of the expected value criterion in the seventeenth century (Hacking 1975: 62). Bernoulli used the St. Petersburg paradox (1738) to refute the expected value concept and then to propose an alternative measure of value based on diminishing marginal returns.[22] This was the first formulation of expected utility, and the concept remained essentially unchanged until it was fully formalized by Von Neumann and Morgenstern (1944). By the 1950s, expected utility theory had gained dominance in economics, but questions about the descriptive accuracy of the theory's axioms and predictions led social psychologists to engage in a series of experiments to see if individuals did in fact behave according to the predictions of expected utility. By the late 1970s, there was growing evidence, primarily from experiments in the laboratory but also from empirical studies of consumer and investment behavior, regarding a number of systematic deviations from expected utility theory.

These were discrete, inductively generated findings generated by dissatisfaction with the descriptive accuracy of expected utility theory, with no apparent connection between those findings. What propelled the research program forward was a series of new conjectures, as economists and social psychologists proposed alternative theories of risky choice by relaxing one or more of axioms of expected utility theory. This lead to a variety of formulations of generalized utility theory (Machina 1982; Camerer 1992). One of the most influential of the alternative theories was prospect theory (Kahneman and Tversky 1979), which emphasized the importance of reference points, the asymmetry of gains and losses around a reference point, and nonlinear responses to probabilities.[23] Prospect theory has been applied in a number of disciplines and has attracted particular attention in international relations (Farnham 1994; McDermott 1998; Levy 2000). It began as a theoretical conjecture in response to a series of apparent experimental and empirical refutations of a prior conjecture about the nature of choice.

The results of ongoing experimental work are mixed. Most analysts agree that there are a number of robust descriptive violations of expected utility, but no single alternative conjecture has replaced it, leaving expected utility theory and prospect theory among a handful of leading contenders to a behavioral theory of choice (Camerer 1992: 239–42). It is important to note that a major reason for the persistence of expected utility theory, in addition to the limitations of competing theories, is its normative appeal as a theory of how people ought to maximize value, even among scholars who are convinced of the descriptive inadequacy of the theory.[24]

The Impact of Policy and Politics[25]

Few would deny that policy and politics often shape the development and persistence of scholarly research programs.[26] Scholars from a variety of metatheoretical orientations have argued that some of the leading research programs in the field are driven by current events and by the policy agendas of states and of individual scholars, that the study of international relations in different countries reflects and therefore varies with their country's distinctive historical circumstances and their government's different policy agendas, and that the relatively new field of international relations reflects a strong American thrust in both policy orientation and academic style (Hoffmann 1977; Krippendorf 1987; Ross 1991; Oren 1995; Waever 1998; Jervis 1998, 2003; Wendt 1999).

In terms of paradigmatic debates, for example, scholars have argued that the interwar period led to idealist and liberal approaches,[27] World War II to realism, the Vietnam War to critical orientations, and the uncertainty of the post–cold war period to multiple paradigms. In terms of substantive focus, the cold war gave rise to an emphasis on nuclear weapons, deterrence theory, and the East-West divide in general. The increase in civil wars and armed insurgencies after the end of the cold war led to a significant expansion of research on ethnonationalism, civil wars, genocide, humanitarian intervention, and, after the September-11 attacks, terrorism.

Similar arguments can be applied to the study of history, where interpretations of the past are often shaped by contemporary values and policy. Combs (1983) argued that changing interpretations of American foreign policy over time reflect ever-changing American foreign policy agendas. The idea that contemporary values, norms, issues, and agendas shape the interpretation of the past is reflected in Croce's famous statement that "all history is contemporary history" (cited in Carr 1964: 20–21), and in Kierkegaard's idea that "life is lived forward but written backwards" (cited in Jervis 2003:100).

While government policy agendas often shape academic research programs—through government or foundation support for academic research or through more diffuse mechanisms—the diversity of the academy reflects a wide range of values and policy agendas, and leading scholarly research programs may reflect agendas and values that are far from the dominant ones in state and society. U.S. government policy agendas shaped traditional histories of the origins of the cold war (Feis 1970), but at the same time competing policy preferences shaped revisionist interpretations of American foreign policy in the 1960s and 1970s (Williams 1972). The influence of countercultural values is also clear in postmodern and cultural history, values that examine the past from the perspective of the powerless and the voiceless and that are currently dominant (and not without power or voice) within many history departments.[28] Perhaps not coincidently, diplomatic and particularly military history, especially in the United States, have been marginalized (Lynn 1997, Black 2004).

Although it is undeniable that politics and policy affect the initiation and evolution of many research programs, it is important to recognize that some influential research programs in the field are driven primarily by autonomous theoretical or analytical developments or by evidentiary support, rather than by recent events or policy agendas. As argued in the last section, the bargaining model of war, rational choice theory in general, and behavioral decision theory have no obvious connection to world events or policy agendas.[29]

Conclusion

I have focused on the descriptive question of what influences the historical evolution of research programs rather than on the more normative questions of how research programs should develop and how they should be evaluated. I have distinguished between theoretical, empirical, and political criteria but conceded that the relationships among them are complex and sometimes difficult to disentangle. My argument is that social science research programs follow multiple trajectories, and that there is no single path for a research program's "success," defined in terms of the program's impact on and endurance in the field. Most rational choice models of international relations are more theory-driven than evidence-driven; though in some cases (Bueno de Mesquita 1981 comes to mind), the ability of some of these models to outperform their rivals in terms of degree of empirical support significantly enhances their influence. Research on the relationship between territory and war and between the dyadic balance of power and the outbreak of war has been primarily evidence-driven. While behavioral decision theory itself has in many respects been evidence-driven, if it is conceived more broadly as part of a broader research program on choice under conditions of risk that goes back to Pascal and Bernoulli, it is a classic case of Popper's model of an alternating sequence of conjectures and refutations.[30]

Policy agendas and normative concerns have had a much greater impact on the evolution of the democratic peace research program, but I question the common view that these factors have been the dominant force behind the scholarly popularity of the democratic peace. I argue instead that the unprecedented levels of empirical support for the dyadic democratic peace proposition, in a field notoriously lacking in law-like behavior, were the primary driving force behind the evolution of the research program, particularly in its early stages.

The question of the relative impact of policy agendas and values on academic research programs is both descriptively interesting and normatively complex.[31] The important question, from the perspective of a normative theory of science, is not *whether* normative and policy concerns influence research programs—since they inevitably do—but *how*. It makes a difference where in the research process normative and policy concerns have an impact. Popper (1965) distinguished between the logic of discovery and the logic of confirmation. The integrity of science is not undermined if values or policy concerns help shape the questions that scholars ask or even the initial theoretical conjectures constructed to explain them. Indeed, social science is a social enterprise as well as a scientific one, and social scientists should be social critics as well as social scientists.[32] As social critics, they should identify and explore important social questions, however out of fashion or contrary to governmental policy they may be.

It is a more serious threat to the integrity of scientific inquiry if values and policy concerns have a significant impact on how scholars define their concepts, translate their conjectures into rigorously formulated theories, construct research designs to test those theories, interpret the evidence,

and decide—in the face of disconfirming evidence—whether or not to abandon the research program. This is not to say that the influence of these factors can be entirely eliminated from these stages of research, but rather that this influence and its negative consequences can be minimized if a scholar acknowledges her underlying normative assumptions and attempts to compensate for them in the construction of her research design.

We should also remember that the inseparability of facts and values is a reciprocal relationship. It means both that normative values infuse all empirical inquiry and that normative arguments have empirical components. Social scientists should be sensitive to the normative assumptions and implications of various theoretical arguments and of the research designs constructed to test them. At the same time, scholars should make a serious effort to identify the empirical components of normative arguments and to test those implied empirical propositions with rigorous social science methods (Snyder 2003).

Notes

1. As Lichbach notes in his concluding essay in this volume, Lakatos has declined in favor in the current literature in the philosophy of science. See also Blaug (1994: 109–11).
2. For the purposes of this study I define research programs broadly to include either 1) a body of scholarship that is built around a well-defined set of theoretical assumptions, which is inherent in Lakatos's (1970) conception, or 2) a body of scholarship that focuses on a well-defined substantive problem. Thus I classify the vast literature on the democratic peace as a research program for the purposes of this study, though Lakatosian criteria would lead us to exclude it because of the variety of theoretical explanations that have been advanced for the democratic peace and the different assumptions on which they are based.
3. Similarly, Jervis (1998: 972) states that "a research program succeeds when many scholars adopt it." The impact of a research program could also be measured in terms of the number of articles in prestigious journals and presses, readings on graduate syllabi, convention panels, doctoral dissertations, and hiring patterns.
4. This typology of theory-driven, evidence driven, and alternating sequence of theory and evidence mirrors Lakatos's (1970: 151–52) conception of three "typical variants" in the evolution of research programs: a "Popperian alternation of conjectures and refutations," a "period of relative autonomy of theoretical progress," and one in which all the empirical evidence is in place prior to

theoretical development. Lakatos suggests that "which pattern is actually realized depends only on historical accident" (p. 151). I thank Mark Lichbach for pointing out that my categories were similar to those of Lakatos.

5. Given the enormous differences in the study of international relations and international history across national boundaries (Smith 1985; Waever 1998; Levy 2001), the inclusion of non-American scholarship would be particularly valuable in isolating the role of politics and policy agendas, which vary across states in a way that theory and evidence presumably do not.
6. Falsifiability is a logical criterion that refers to whether the theory or hypothesis is constructed in such a way that there is a nonempty set of empirical observations that would lead researchers to conclude that the theory was incorrect, or at least that it needed to be rejected. Whether a theory is actually falsified is an empirical question, though one that involves some difficult issues in the philosophy of science. See the Lichbach chapter in this volume.
7. A theory is parsimonious if it explains as much as possible with as little theoretical apparatus as possible. A theory is not parsimonious in the abstract but only relative to other theories that purport to explain the same phenomenon. In this view parsimony relates to theories that one constructs to explain the world, not to beliefs about the simplicity of the world itself. King, Keohane, and Verba (1994) adopt this second definition of parsimony, and refer to the first as "maximizing leverage."
8. One of the problems with Lakatos's (1970) conception of research programs is the ambiguity surrounding the "unit of appraisal," or how broadly one should define research programs (DiCicco and Levy 2003). Should we focus, for example, on the rational choice paradigm as a whole; on a particular analytic framework within that paradigm, such as games of incomplete information or, more narrowly, signaling game models; or on applications of rational choice to a particular substantive area, such as bargaining?
9. This discussion builds on Levy (2003b). For a broader analysis of bargaining, one that includes nonrational factors, see Lebow (1996).
10. The key analytic developments were the treatment of games of incomplete information (about adversary preferences) as games of imperfect information (about prior moves in the game) (Harsanyi 1967–68), and the refinement of key equilibrium concepts that permitted the solution of these games. The key equilibrium concepts include perfect equilibrium (Selton 1975) and sequential equilibrium (Kreps and Wilson 1982). Rubenstein (1982) first applied perfect equilibrium to bargaining problems.
11. Within realism, for example, one can identify classical realism and structural realism, offensive realism and defensive realism (Walt 2002), balance of power realism and hegemonic realism (Levy 2002). This leads us back to the question of the appropriate "unit of appraisal," and the question of whether these broad paradigms are usually conceived as a single integrated research program.
12. This is somewhat ironic, because many of the scholarly contributions of King, and especially of Keohane and Verba in their individual work have been more theoretical than empirical.

13. Another example of a research program—or perhaps a paradigmatic approach—that is driven more by evidence than by theory is cognitive psychology. I thank Ned Lebow for suggesting this example.
14. I define the "democratic peace" research program broadly here to include not only research related to the dyadic-level proposition that democracies rarely if ever fight each other, but also monadic-level propositions about the relative war-proneness of democratic states. For an intellectual history of the research program see Ray (1995).
15. On the early skepticism of these scholars toward the democratic peace see Ray (1995: 44).
16. On the progressive nature of the democratic peace research program, as judged by several alternative metatheoretical criteria, see Chernoff (2005).
17. The problem is compounded by confusion over levels of analysis. Balance of power theory and power transition theories are system level, while the power parity and power preponderance hypotheses are dyadic level.
18. For an early review see Siverson and Sullivan (1983).
19. The dominant role of theory in the debate between balance of power theory and power transition theory derives in part from a certain amount of incomensurability between the two. Most balance of power theories focus on land-based military power and are applied to continental systems, especially Europe, while most hegemonic theories, including power transition theory, emphasize economic foundations of power and are applied to global maritime systems (Levy 2003a).
20. On the limitations of physics as a model for the social sciences, and for the possible relevance of other natural sciences, including biology, see Bernstein et al. (2000, and in this volume). On the relevance of other disciplines for the study of history, see Gaddis (2002).
21. Some of the key studies in behavioral decision theory can be found in Edwards and Tversky (1967), Kahneman and Tversky (2000), Kagel and Roth (1995), and Camerer, Loewenstein, and Rabin (2004).
22. This was a theoretical refutation based on the identification of a theoretical anomaly in the expected value concept.
23. Prospect theory itself has been significantly revised (Tversky and Kahneman 1992) in response to an important theoretical (as opposed to empirical) problem relating to the mathematical intractability of the original formulation of the probability weighting function.
24. The normative appeal of expected utility theory implies no distinctive substantive commitment, other than the maximization of individual value, independently of how the individual defines value. Rational choice theory includes "analytical Marxists" such as Przeworski as well as free market economists. Prospect theory, on the other hand, makes no normative claims, and Kahneman and Tversky (1979) and others argue that it is impossible to reconcile normative and descriptive theories of choice.
25. This section builds on Levy (2003b).

26. "Policy and politics" is a very broad category, and includes the impact of current events, the policy agendas of the government and of political oppositions or other groups. It might also include the professional or financial self-interest of individual scholars or research teams, but I exclude this latter consideration from this discussion.
27. Carr (1939) argued that the ascendance of idealist international theory, with its vision of a natural harmony of interests in the world, was basically a rationalization for British and American dominance in a liberal world order.
28. For an argument on why the study of the "voiceless" lends itself to a postmodern orientation, see Haber, Kennedy, and Krasner (1997): 38–40.
29. The rational choice paradigm does not specify actors values or preferences, which are exogenous. It specifies how actors should behave, and perhaps how they do behave, given their values, their beliefs, and the structure of their structural and informational environments. Moreover, the actors themselves are unspecified. They can be individuals, organizations, classes, states, empires, intergovernmental organizations, or any group whose preferences satisfy the axioms of expected utility theory, or perhaps even less demanding criteria in "softer" versions of rational choice. Particular rational choice theories (the signaling model of economic interdependence and peace, for example) specify actors, preferences, and other parameters.
30. This suggests that the temporal boundaries we ascribe to a research program may affect how we classify it. Behavioral decision research from the 1950s to the late 1970s was primarily empirically driven, but if we focus on decision theory more broadly to include the antecedents of behavioral decision research and the theories it generated, it fits nicely into a model of alternating conjectures and refutations.
31. For a discussion of the tension between prescriptive and descriptive theories of research programs, between the methodology of science and the history of science, see Blaug (1994).
32. This does not imply that all researchers need to devote equal time and energy to social criticism and scientific analysis, only that both tasks are appropriate ones.

References

Arrow, Kenneth J. 1951. *Social Choice and Individual Values*. New York: Wiley.

Backhouse, Roger E. 1994. "The Lakatosian Legacy in Economic Methodology," in Backhouse, ed., *New Directions in Economic Methodology*. London: Routledge, pp. 173–91.

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast. 1998. *Analytic Narratives*. Princeton, NJ: Princeton University Press.

Bernstein, Steven, Richard Ned Lebow, Janice Gross Stein, and Steven Weber. 2000. "God Gave Physics the Easy Problems: Adapting Social Science to an Unpredictable World." *European Journal of International Relations*, Vol. 6, No. 1: 43–96.

Black, Jeremy. 2004. *Rethinking Military History*. London: Routledge.

Blainey, Geoffrey. 1973. *The Causes of War*. New York: Free Press. Chap. 8.

Blaug, Mark. 1994. "Why I am Not a Constructivist: Confessions of an Unrepentant Popperian," in Roger E. Backhouse, ed., *New Directions in Economic Methodology*. New York: Routledge, pp. 109–36.

Brady, Henry E., and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Bremer, Stuart A. 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816–1965." *Journal of Conflict Resolution*, Vol. 36, No. 2 (June): 309–41.

Bueno de Mesquita. 1981. *The War Trap*. New Haven, CT: Yale University Press.

Bueno de Mesquita, Bruce, and David Lalman. 1992. *War and Reason*. New Haven, CT: Yale University Press.

Bueno de Mesquita, Bruce, James D. Morrow, Randolph M. Siverson, and Alastair Smith. 2003. *The Logic of Political Survival*. Cambridge, MA: MIT Press.

Bull, Hedley. 1966. "International Theory: The Case for a Classical Approach." *World Politics*, Vol. 18, No. 3 (April): 361–77.

Camerer, Colin F. 1992. "Recent Tests of Generalizations of Expected Utility Theory," in Ward Edwards, ed., *Utilities: Measurement and Applications*. Boston, MA: Kluwer Academic Publishers.

Camerer, Colin F., George Loewenstein, and Matthew Rabin, eds. 2004. *Advances in Behavioral Economics*. New York and Princeton, NJ: Russell Sage and Princeton University Press.

Carr, Edward Hallet. 1939. *The Twenty Years' Crisis*. New York: Harper Torchbooks.

———. 1964. *What Is History?* Harmondsworth, UK: Penguin Books.

Chernoff, Fred. 2005. *The Power of International Theory: Reforging The Link To Foreign Policy-Making Through Scientific Enquiry*. London: Routledge.

Combs, Jerald. 1983. *American Diplomatic History: Two Centuries of Changing Interpretations*. Berkeley, CA: University of California Press.

DiCicco, Jonathan M., and Jack S. Levy. 2003. "The Power Transition Research Program: Theoretical Development, Empirical Corroboration, and Metatheoretical Implications," in Colin Elman and Miriam Fendius Elman, eds.,*Progress in International Relations Theory: Appraising the Field*. Cambridge, MA: MIT Press, pp. 109–57.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: HarperCollins.

Doyle, Michael W. 1983. "Kant, Liberal Legacies, and Foreign Affairs: Parts I & II." *Philosophy and Public Affairs*, Vol. 12: 205–35, 323–53.

Edwards, Ward, and Amos Tversky, eds. 1967. *Decision Making*. Baltimore, MD: Penguin Books.

Elman, Colin, and Miriam Fendius Elman, eds. 2003. *Progress in International Relations Theory: Appraising the Field*. Cambridge, MA: MIT Press.

Farnham, Barbara 1994. *Taking Risks/Avoiding Losses*. Ann Arbor, MI: University of Michigan Press.
Fearon, James. 1995. "Rationalist Explanations for War." *International Organization*, Vol. 9, No. 3 (Summer): 379–414.
Fearon, James D., and David D. Laitin. 1996. "Explaining Interethnic Cooperation." *American Political Science Review*, Vol. 90, No. 4 (December): 715–35.
Feis, Herbert. 1970. *From Trust to Terror: The Onset of the Cold War 1945–1950*. New York: W.W. Norton and Company.
Filson, Darren, and Suzanne Werner. 2002. "A Bargaining Model of War and Peace: Anticipating the Onset, Duration, and Outcome of War." *American Journal of Political Science*, Vol. 46 (October): 819–38.
Gaddis, John Lewis. 2002. *The Landscapes of History: How Historians Map the Past*. Oxford: Oxford University Press.
Gartzke, Eric. 1999. "War Is in the Error Term." *International Organization*, Vol. 53, No. 3 (Summer): 567–87.
George, Alexander L., and Andrew Bennett (2005) *Case Studies and Theory Development*. Cambridge, MA: MIT Press.
Gerring, John. 2004. "What Is a Case Study and What Is It Good for?" *American Political Science Review*, Vol. 98, No. 2 (May): 341–54.
Gilovich, Thomas, Dale Griffin, and Daniel Kahneman, eds. 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press.
Gilpin, Robert. 1975. *U.S. Power and the Multinational Corporation*. New York: Basic Books.
Gowa, Joanne. 1999. *Ballots and Bullets*. Princeton, NJ: Princeton University Press.
Grether, D.M., and C.R. Plott. 1979. "Economic Theory of Choice and the Preference Reversal Phenomenon." *American Economic Review*, Vol. 69: 623–38.
Guzzini, Stefano. 1998. *Realism in International Relations and International Political Economy: The Continuing Story of a Death Foretold*. London: Routledge.
Haber, Stephen H., David M. Kennedy, and Stephen D. Krasner. 1997. "Brothers under the Skin," *International Security*, Vol. 22, No. 1 (Summer): 33–43.
Hacking, Ian. 1975. *The Emergence of Probability*. Cambridge, England: Cambridge University Press.
Harsanyi, John C. 1967–68. "Games with Incomplete Information Played by Bayesian Players," Parts 1–3. *Management Science*, Vol. 14: 159–82, 320–34, 486–502.
Hempel, Carl G. 1966. *Philosophy of Science*. Englewood Cliffs, NJ: Prentice-Hall.
Hensel, Paul R. 2000. "Theory and Evidence on Geography and Conflict," in John A. Vasquez, ed., *What Do We Know About War?* Lanham, MD: Rowman & Littlefield Publishers, Inc., pp. 57–84.
Hoffmann, Stanley. 1977. "An American Social Science: International Relations." *Daedalus* (Summer): 41–59.
Holsti, K.J. 1985. *The Dividing Discipline: Hegemony and Diversity in International Theory*. Boston, MA: Allen & Unwin.

Huth, Paul K. 1996. *Standing Your Guard: Territorial Disputes and International Conflict*. Ann Arbor, MI: University of Michigan Press.
Huth, Paul K., and Todd L. Alee. 2002. *The Democratic Peace and Territorial Conflict in the 20th Century*. New York: Cambridge University Press.
Iggers, Georg G. 1984. *New Directions in European Historiography*, rev. ed. Middletown, CT: Wesleyan University Press.
Jervis, Robert. 1998. "Realism and the Study of World Politics." *International Organization*, Vol. 52, No. 4 (Autumn): 971–91.
———. 2003. "Security Studies: Ideas, Policy, and Politics," in Edward D. Mansfield and Richard Sisson, eds., *The Evolution of Political Knowledge: Democracy, Autonomy, and Conflict in Comparative and International Politics*. Columbus: Ohio State University Press, pp. 100–26.
Kagel, John H., and Alvin E. Roth, eds. 1995. *The Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.
Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, Vol. 47: 263–91.
———, eds. 2000. *Choices, Values, and Frames*. New York: Cambridge University Press.
Kaplan, Morton. 1966. "The New Great Debate: Traditionalism vs. Science in International Relations." *World Politics*, Vol. 19, No. 1 (October): 1–20.
King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
Kreps, David M., and Robert Wilson. 1982. "Sequential Equilibria." *Econometrica*, Vol. 50: 863–94.
Krippendorf, Ekkehart. 1987. "The Dominance of American Approaches in International Relations," *Millennium*, Vol. 16, No. 2 (Summer): 207–14.
Kugler, Jacek, and Douglas Lemke, eds. 1996. *Parity and War*. Ann Arbor, MI: University of Michigan Press.
Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes," in I. Lakatos and Alan Musgrave, eds., *Criticism and the Growth of Knowledge*. New York: Cambridge University Press, pp. 91–196.
Lake, David A. Lake and Donald Rothchild, eds. 1996. *The International Spread of Ethnic Conflict*. Princeton, NJ: Princeton University Press.
Lapid, Yosef. 1989. "The Third Debate: On the Prospects of International Theory in a Post-Positivist Era." *International Studies Quarterly*, Vol. 33, No. 3 (September): 235–54.
Laudan, Larry. 1978. *Progress and Its Problems: Toward a Theory of Scientific Growth*. Berkeley, CA: University of California Press.
Lebow, Richard Ned. 1996. *The Art of Bargaining*. John Hopkins University Press.
Levy, Jack S. 2000. "Loss Aversion, Framing Effects, and International Conflict," in Manus I. Midlarsky, ed., *Handbook of War Studies II*. Ann Arbor, MI: University of Michigan Press, pp. 193–221.
———. 2001. "Explaining Events and Testing Theories: History, Political Science, and the Analysis of International Relations," in Colin Elman and Miriam

Fendius Elman, eds., *Bridges and Boundaries: Historians, Political Scientists, and the Study of International Relations*. Cambridge, MA: MIT Press, pp. 39–83.

Levy, Jack S. 2002. "War and Peace," in Walter Carlsnaes, Thomas Risse, and Beth A. Simmons, eds., *Handbook of International Relations*. London: Sage Publications, pp. 350–68.

———. 2003a. "Balances and Balancing: Concepts, Propositions, and Research Design," in John A. Vasquez and Colin Elman, eds., *Realism and the Balancing of Power: A New Debate*. Englewood Cliffs, NJ: Prentice-Hall, pp. 128–53.

———. 2003b. "Policy and Politics in Security Studies," in Edward D. Mansfield and Richard Sisson, eds., *The Evolution of Political Knowledge: Democracy, Autonomy, and Conflict in Comparative and International Politics*. Columbus: Ohio State University Press, pp. 127–30.

Lynn, John A. 1997. "The Embattled Future of Academic Military History," *Journal of Military History*, Vol. 61, No. 4 (October): 777–89.

Machina, Mark J. 1982. " 'Expected Utility' Analysis without the Independence Axiom." *Econometrica*, Vol. 50: 277–323.

Maoz, Zeev. 1997. "The Controversy over the Democratic Peace: Rearguard Action or Cracks in the Wall?" *International Security*, Vol. 22 (Summer): 162–98.

McDermott, Rose. 1998. *Risk-Taking in International Politics: Prospect Theory in American Foreign Policy*. Ann Arbor, MI: University of Michigan Press.

Merton, Robert K. 1967. "The Bearing of Sociological Theory on Empirical Research," in Merton, *On Theoretical Sociology*. New York: Free Press. Chap. IV.

Morgenthau, Hans. 1948. *Politics Among Nations: The Struggle for Power and Peace*. New York: Knopf.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.

Oren, Ido. 1995. "The Subjectivity of the Democratic Peace: Changing U.S. Perceptions of Imperial Germany." *International Security*, Vol. 20, No. 2 (Fall): 147–84.

Organski, A.F.K. 1968. *World Politics*, 2nd ed. New York: Knopf.

Popper, Karl R. 1965. *The Logic of Scientific Discovery*. New York: Harper Torchbooks.

———. 1962. *Conjectures and Refutations*. New York: Basic Books.

Powell, Robert. 2002. "Bargaining Theory and International Conflict." *Annual Review of Political Science*, Vol. 5: 1–30.

Ray, James Lee. 1995. *Democracy and International Politics: An Evaluation of the Democratic Peace Proposition*. Columbia, SC: Columbia: University of South Carolina Press.

Ross, Dorothy. 1991. *The Origins of American Social Science*. New York: Cambridge University Press.

Rubenstein, Ariel.1982. "Perfect Equilibrium in a Bargaining Model." *Econometrica*, Vol. 50, No. 1: 97–110.

Ruggie, John Gerard. 1998. "What Makes the World Hang Together: Neo-Utilitarianism and the Social Constructivist Challenge." *International Organization*, Vol. 52, No. 4 (Autumn): 855–85.

Russett, Bruce. 1993. *Grasping the Democratic Peace*. Princeton, NJ: Princeton University Press.

Schmidt, Brian C. 1998. *The Political Discourse of Anarchy: A Disciplinary History of International Relations*. Albany: State University of New York Press.

Selton, Reinhart. 1975. "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory*, Vol. 4: 25–55.

Siverson, Randolph M., and Michael P. Sullivan. 1983. "The Distribution of Power and the Onset of War." *Journal of Conflict Resolution*, Vol. 27: 473–94.

Slantchev, Branislav L. 2003. "The Principle of Convergence in Wartime Negotiations." *American Political Science Review*, Vol. 97, No. 4 (November): 621–32.

Small, Melvin, and J. David Singer. 1976. "The War-Proneness of Democratic Regimes, 1816–1965." *Jerusalem Journal of International Relations*, Vol. 1 (Summer): 50–69.

Smith, Steve. 1985. *International Relations: British and American Perspectives*. Oxford: Blackwell.

Snyder, Jack. 2003. " 'Is' and 'Ought': Evaluating Empirical Aspects of Normative Research," in Colin Elman and Miriam Fendius Elman, eds., *Progress in International Relations Theory: Appraising the Field*. Cambridge, MA: MIT Press, pp. 349–77.

Vasquez, John A. 1983. *The Power of Power Politics: A Critique*. New Brunswick, NJ: Rutgers University Press.

———. 1993. *The War Puzzle*. New York: Cambridge University Press.

Vasquez, John A., and Colin Elman, eds. 2003. *Realism and the Balancing of Power: A New Debate*. Upper Saddle River, NJ: Prentice-Hall.

Vasquez, John A., and Paul D. Senese. 2004. "Alliances, Territorial Disputes, and the Probability of War: Testing for Interactions," in Paul F. Diehl, ed., *The Scourge of War*. Ann Arbor, MI: University of Michigan Press, pp. 189–221.

Von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Waever, Ole. 1998. "The Sociology of a Not So International Discipline: American and European Developments in International Relations." *International Organization*, Vol. 52, No. 4 (Autumn): 687–727.

Wagner, R. Harrison. 2000. "Bargaining and War." *American Journal of Political Science*, Vol. 44: 469–85.

Walt, Stephen M. 2002. "The Enduring Relevance of the Realist Tradition," in Ira Katznelson and Helen V. Milner, *Political Science: State of the Discipline*. New York: W.W. Norton and Company, pp. 197–230.

Waltz, Kenneth N. 1997. "Evaluating Theories," *American Political Science Review*, Vol. 91, No. 4 (December): 913–17.

Weede, Erich. 1984. "Democracy and War Involvement." *Journal of Conflict Resolution*, Vol. 28, No. 4 (December): 649–64.

Weintraub, E. Roy. 1985. *General Equilibrium Analysis: Studies in Appraisal* (New York: Cambridge University Press.

Williams, William Appleman. 1972. *The Tragedy of American Diplomacy*, 2nd ed. New York: Dell.