The Realist Paradigm and Degenerative versus Progressive Research Programs: An Appraisal of Neotraditional Research on Waltz's Balancing Proposition

JOHN A. VASQUEZ *Vanderbilt University*

Several analysts argue that, despite anomalies, the realist paradigm is dominant because it is more fertile than its rivals. While the ability of the realist paradigm to reformulate its theories in light of criticism accounts for its persistence, it is argued that the proliferation of emendations exposes a degenerating tendency in the paradigm's research program. This article applies Lakatos's criterion that a series of related theories must produce problemshifts that are progressive rather than degenerating to appraise the adequacy of realist-based theories on the balancing of power advanced by neotraditionalists. This research program is seen as degenerating because of (1) the protean character of its theoretical development, (2) an unwillingness to specify what constitutes the true theory, which if falsified would lead to a rejection of the paradigm, (3) a continual adoption of auxiliary propositions to explain away flaws, and (4) a dearth of strong research findings.

Within international relations inquiry, the debate over the adequacy of the realist paradigm has been fairly extensive since the 1970s. In Europe it is often referred to as the interparadigm debate (see Banks 1985; Smith 1995, 18–21). In North America, the focus has been more singularly on realist approaches and their critics (see Vasquez 1983). Toward the end of the 1970s, it appeared that alternate approaches, such as transnational relations and world society perspectives, would supplant the realist paradigm. This did not happen, partly because of the rise of neorealism, especially as embodied in the work of Waltz (1979). Now the debate over the adequacy of the realist paradigm has emerged anew.

In this analysis, *realism* is defined as a set of theories associated with a group of thinkers who emerged just before World War II and who distinguished themselves from idealists (i.e., Wilsonians) on the basis of their belief in the centrality of power for shaping politics, the prevalence of the practices of power politics, and the danger of basing foreign policy on morality or reason rather than interest and power. The *realist paradigm* refers to the shared fundamental assumptions various realist theorists make about the world. Derived primarily from the exemplar of realist scholarship, Morgenthau's ([1948] 1978) *Politics among Nations*, these include: (1) Nation-states are the most important actors in international relations; (2) there is a sharp distinction between domestic and international politics; and (3) international relations is a struggle for power and peace. Understanding how and why that struggle occurs is the major purpose of the discipline (see Vasquez 1983, 15–9, 26–30 for elaboration and justification).

While much of the debate over realism has focused on a comparison to neoliberalism (see Kegley 1995),[1] the debate has also raised new empirical (Rosecrance and Stein 1993), conceptual (Lebow and Risse-Kappen 1995, Wendt 1992), and historical (Schroeder 1994a) challenges to the paradigm as a whole. Some call for a sharp break with the paradigm (e.g., Vasquez 1992), while others see the need to reformulate on the basis of known empirical regularities (Wayman and Diehl 1994). Many still see it as the major theoretical framework within which the field must continue to work (Hollis and Smith 1990, 66), and even critics like Keohane ([1983] 1989) and Nye (1988) see the need to synthesize their approaches (in this case neoliberalism) with the realist paradigm.

If any progress is to be made, scholars must have a set of criteria for appraising the empirical component of theories and paradigms (see Vasquez 1992, 1995). Appraising a paradigm, however, is difficult because often its assumptions are not testable, since they typically do not explain anything in and of themselves (e.g., nation-states are the most important actors). Essentially, a paradigm promises scholars that if they view the world in a particular way, they will successfully understand the subject they are studying. In Kuhn's ([1962] 1970, 23–4) language, paradigms do not so much provide answers as the promise of answers. Ultimately, a paradigm must be appraised by its utility and its ability to make good on its promise. Thus, a paradigm can only be appraised indirectly by examining the ability of the theories it generates to satisfy criteria of adequacy.

Within mainstream international relations, the work of Lakatos (1970) has attracted the most consensus as a source of such criteria among both quantitative and

John A. Vasquez is Professor of Political Science, Vanderbilt University, Nashville, TN 37235.

The author thanks Marie T. Henehan and the anonymous reviewers for helpful comments and suggestions.

[1] "Neoliberalism" is a label employed by a number of scholars (see Nye 1988, 1993, 36–40) to refer to a theoretical approach associated with a cluster of three ideas: (1) Democracies do not fight one another (an idea going back at least to Kant); (2) free trade and growing wealth will create a harmony of interests that will reduce the need for war (the position of the early free traders); and (3) reason can be used to reduce global anarchy and produce more orderly relations among states in part through the creation of global institutions (ideas associated with Grotius and, later, Wilson). For a complete review, see the authors in Kegley 1995; see also Doyle 1986.

traditional scholars (see Keohane [1983] 1989). Although the appraisal of theories and the paradigms from which they are derived involves a number of criteria (see Simowitz and Price 1990), including, in particular, the criterion of empirical accuracy (the ability to pass tests) and the principle of falsifiability, the present analysis will apply only the main criterion on which Lakatos laid great stress for the evaluation of a series of theories: They must produce a progressive as opposed to a degenerating research program. Lakatos's criteria clearly stem from a more positivist perspective, but since realists and neorealists accept them, they are perfectly applicable.[2]

One main difference between Lakatos and early positivists is that Lakatos believes the rules of theory appraisal are community norms and cannot be seen as logically compelling, as Popper (1959) had hoped. The case that any given research program is degenerating (or progressive) cannot be logically proven. Such a stance assumes a foundationalist philosophy of inquiry that has been increasingly under attack in the last two decades (see Hollis and Smith 1990). A more reasonable stance is that exemplified by the trade-off between type 1 and type 2 errors in deciding to accept or reject the null hypothesis. Deciding whether a research program is degenerating involves many individual decisions about where scholars are willing to place their research bets, as well as collective decisions as to which research programs deserve continued funding, publication, and so forth. Some individuals will be willing to take more risks than others. This analysis seeks to present evidence that is relevant to the making of such decisions.

The task of determining whether research programs are progressive or degenerating is of especial importance because a number of analysts (e.g. Hollis and Smith 1990, 66; Wayman and Diehl 1994, 263) argue that, despite anomalies, the realist paradigm is dominant because it is more enlightening and fertile than its rivals. While the ability of the realist paradigm to reformulate its theories in light of conceptual criticism and unexpected events is taken by the above authors as an indicator of its fertility and accounts for its persistence, the proliferation of emendations may not be a healthy sign. Indeed, it can be argued that persistent emendation exposes the degenerating character of the paradigm. This analysis will demonstrate that the "theoretical fertility" apparently exhibited by realism in the last twenty years or so is actually an indicator of the degenerating nature of its research program.

THE CRITERION

Imre Lakatos (1970) argued against Popper (1959) and in favor of Kuhn ([1962] 1970) that no single theory can ever be falsified because auxiliary propositions can be added to account for discrepant evidence. The problem, then, is how to evaluate a *series of theories* that are intellectually related.

A series of theories is exactly what is posing under the general rubrics of realism and neorealism. All these theories share certain fundamental assumptions about how the world works.[3] In Kuhn's ([1962] 1970) language, they constitute a family of theories because they share a paradigm. A *paradigm* can be stipulatively defined as "the fundamental assumptions scholars make about the world they are studying" (Vasquez 1983, 5).[4] Since a paradigm can easily generate a family of theories, Popper's (1959) falsification strategy was seen by Lakatos (1970) as problematic, since one theory can simply be replaced by another in incremental fashion without ever rejecting the shared fundamental assumptions. It was because of this problem that Kuhn's sociological explanation of theoretical change within science was viewed as undermining the standard view in philosophy of science, and it was against Kuhn that Lakatos developed his criteria for appraising a series of theories. To deal with the problem of appraising a series of theories that may share a common paradigm or set of assumptions, Lakatos stipulated that a research program coming out of this core must develop in such a way that theoretical emendations are progressive rather than degenerating.

The main problem with this criterion is that, unless it is applied rigorously, with specific indicators as to what constitutes "progressive" or "degenerating" research programs, it will not provide a basis for settling the debate on the adequacy of the realist paradigm. In an early application of this criterion to structural realism, Keohane ([1983] 1989, 43–4, 52, 55–6, 59), for example, goes back and forth talking about not only the fruitfulness of neorealism but also its incompleteness and the general inability of any international relations theory to satisfy Lakatos's criteria (see also Nye 1988, 243).

Eventually, it would be highly desirable to construct operational indicators of the progressive or degenerating nature of a paradigm's research program. Since these are not available, this analysis will explicitly identify the characteristics that will be used to indicate that a research program is degenerating. Lakatos (1970, 116–7) sees a research program as degenerating if its auxiliary propositions increasingly take on the characteristic of ad hoc explanations that do not produce any novel (theoretical) facts, as well as new empirical content. For Lakatos (p. 116), "no experimental result can ever kill a theory: any theory can be saved from counterinstances either by some auxiliary hypothesis or by a suitable reinterpretation of its terms." Since Lakatos (p. 117) finds this to be the case, he asks: Why not "impose certain standards on the theoretical adjustments by which one is allowed to save a theory?" Adjustments that are acceptable he labels

[2] Vasquez (1995) deals with antifoundationalist postpositivist criticisms of such criteria. On the latter, see Lapid (1989).

[3] Theory is defined here as a set of interrelated propositions purporting to explain behavior; see Vasquez 1992, 835–6. Given this definition, which is noncontroversial, the realist paradigm can have many different theories; see Vasquez 1983, 4–6.

[4] Masterman (1970) has criticized Kuhn for using the concept of paradigm ambiguously. This stipulative definition is meant to overcome this objection, while still capturing the essence of what Kuhn ([1968] 1970, Postscript) was trying to do with the concept.

progressive, and those that are not he labels degenerating.

The key for Lakatos is to evaluate not a single theory but a series of theories linked together. Is each "theoryshift" advancing knowledge, or is it simply a "linguistic device" for saving a theoretical approach?[5] A theoryshift or problemshift is considered (1) theoretically progressive if it theoretically "predicts some novel, hitherto unexpected fact" and (2) empirically progressive if these new predictions are actually corroborated, giving the new theory an excess empirical content over its rival (Lakatos 1970, 118). In order to be considered progressive, a problemshift must be *both* theoretically and empirically progressive—anything short of that is defined (by default) as *degenerating* (p. 118). A degenerating problemshift or research program, then, is characterized by the use of semantic devices that hide the actual content-decreasing nature of the research program through reinterpretation (p. 119). In this way, the new theory or set of theories are really ad hoc explanations intended to save the theory (p. 117).

It should be clear from this inspection of Lakatos's criterion that progressive research programs are evaluated ultimately on the basis of a criterion of accuracy, in that the new explanations must pass empirical testing. If this is the case, then they must in principle be *falsifiable*. The generation of new insights and the ability to produce a number of research tests, consequently, are not indicators of a progressive research program, if *these do not result in new empirical content that has passed empirical tests*.

How can one tell whether a series of theories that come out of a research program is degenerating? First, the movement from T to T' *may* indicate a degenerating tendency if the revision of T involves primarily the introduction of new concepts or some other reformulation that attempts to explain away discrepant evidence. Second, this will be seen as degenerating if this reformulating never points to any novel unexpected facts, by which Lakatos means that T' should tell scholars something about the world other than what was uncovered by the discrepant evidence. Third, if T' does not have any of its new propositions successfully tested or lacks new propositions (other than those offered to explain away discrepant evidence), then it does not have excess empirical content over T, and this can be an indicator of a degenerating tendency in the research program. Fourth, if a research program goes through a number of theoryshifts, all of which have one or more of the above characteristics *and* the end result of these theoryshifts is that collectively the family of theories fields a set of contradictory hypotheses which greatly increase the probability of at least one passing an empirical test, then a research program can be appraised as degenerating.

[5] Lakatos (1970, 118 n3) notes that by "problemshift" he really means "theoryshift" (i.e., a shift from one specific theory to another) but does not use that word because it "sounds dreadful." Actually, it is much clearer. On the claim that the problemshifts which are degenerating are really just linguistic devices to resolve anomalies in a semantic manner, see Lakatos 1970, 117, 119.

This fourth indicator is crucial and deserves greater explication. It implies that while some latitude may be permitted for the development of ad hoc explanations, the longer this goes on in the face of discrepant evidence, the greater is the likelihood that scientists are engaged in a research program that is constantly repairing one flawed theory after another without any incremental advancement in the empirical content of these theories. What changes is not what is known about the world, but semantic labels to describe discrepant evidence that the original theory(ies) did not anticipate.

How does one determine whether semantic changes are of this sort or the product of a fruitful theoretical development and new insights? An effect of repeated semantic changes which are not progressive is that they focus almost entirely on trying to deal with experimental outcomes or empirical patterns contrary to the initial predictions of the theory. One consequence is that collectively the paradigm begins to embody contradictory propositions, such as (1) war is likely when power is not balanced and one side is preponderant, and (2) war is likely when power is relatively equal. The development of two or more contradictory propositions increases the probability that at least one of them will pass an empirical test. If a series of theories, all derived from the same paradigm (and claiming a family resemblance, such as by using the same name, e.g., Freudian, Marxist, or realist), predict several competing outcomes as providing support for the paradigm, then this is an example of the fourth indicator. Carried to an extreme, the paradigm could prevent any kind of falsification, because collectively its propositions in effect pose the bet: "Heads, I win; tails, you lose." A research program can be considered blatantly degenerative if one or more of the behaviors predicted is only predicted after the fact.

To be progressive, a theoryshift needs to do more than just explain away the discrepant evidence. It should show how the logic of the original or reformulated theory can account for the discrepant evidence and then delineate how this theoretic can give rise to new propositions and predictions (or observations) that the original theory did not anticipate. The generation of new predictions is necessary because one cannot logically test a theory on the basis of the discrepant evidence that led to the theoryshift in the first place, since the outcome of the statistical test is already known (and therefore cannot be objectively predicted before the fact). The stipulation of new hypotheses that pass empirical testing on some basis other than the discrepant evidence is the minimal logical condition for being progressive. Just *how* fruitful or progressive a theoryshift is, beyond the minimal condition, depends very much on how insightful and/or unexpected the novel facts embodied in the auxiliary hypotheses are deemed to be by scholars within the field. Do they tell scholars things they did not (theoretically) know before?

It should be clear that the criteria of adequacy involve the application of disciplinary norms as to what constitutes progress. The four indicators outlined above provide reasonable and fairly explicit ways to interpret the evidence. Applying them to a body of research should permit a basis for determining whether a research program appears to be on the whole degenerative or progressive.

It will be argued that what some see as theoretical enrichment of the realist paradigm is actually a proliferation of emendations that prevent it from being falsified. It will be shown that the realist paradigm has exhibited (1) a protean character in its theoretical development, which plays into (2) an unwillingness to specify what form(s) of the theory constitutes the true theory, which if falsified would lead to a rejection of the paradigm, as well as (3) a continual and persistent adoption of auxiliary propositions to explain away empirical and theoretical flaws that greatly exceed the ability of researchers to test the propositions and (4) a general dearth of strong empirical findings. Each of these four characteristics can be seen as "the facts" that need to be established or denied to make a decision about whether a given research program is degenerating.

THE RESEARCH PROGRAM TO BE ANALYZED

Any paradigm worth its salt will have more than one ongoing research program, so in assessing research programs it is important to select those that focus on a core area of the paradigm and not on areas that are more peripheral or can be easily accommodated by a competing paradigm. It also is important that the research program be fairly well developed both in terms of the number of scholars and the amount of time spent on the program.

If one uses Kuhn's ([1962] 1970) analysis to understand the post–World War II development of the field of international relations, there is a general consensus that the realist paradigm has dominated international relations inquiry within the English-speaking world and that Morgenthau's *Politics among Nations* can be seen as the exemplar of this paradigm (see Vasquez 1983 for a test of this claim; see also Banks 1985; Smith 1995; Olson and Groom 1991; and George 1994). Neorealism can be seen as a further articulation of the realist paradigm along at least two lines. The first, by Waltz (1979), brought the insights of structuralism to bear on realism and for this reason is often referred to as structural realism. For Waltz (1979), structure (specifically the anarchic nature of the international system) is presented as the single most important factor affecting all other behavior. The second by Gilpin (1981), brought to bear some of the insights of political economy with emphasis on the effect of the rise and decline of hegemons on historical change. Both of these efforts have developed research programs. Generally, it is fair to say that Waltz has had more influence on security studies, whereas Gilpin has been primarily influential on questions of international political economy. Since the main concern here is with security, peace, and war, this appraisal will concentrate on the work of scholars who have been influenced by Waltz.

A complete case against the realist paradigm needs to look at other aspects of neorealism and to examine classical realism as well. Elsewhere, the quantitative work guided by classical realism has been evaluated (Vasquez 1983). Gilpin's work on war is best treated in conjunction with the power transition thesis of Organski and Kugler (1980), with which it shares a number of similarities (for an initial appraisal see Vasquez 1993, chapter 3; 1996). So, part of the reason for focusing on Waltz and the research agenda sparked by his analysis is that only so much work can be reviewed in depth in a single article.[6] The more compelling reason is that Waltz's analysis has in fact had a great impact on empirical research. His influence on those who study security questions within international relations in what may be called a neotraditional (i.e., nonquantitative) manner is without equal.

Waltz (1979) centers on two empirical questions: (1) explaining what he considers a fundamental law of international politics, the balancing of power, and (2) delineating the differing effects of bipolarity and multipolarity on system stability. While the latter has recently given rise to some vehement debates about the future of the post–Cold War era (see Mearsheimer 1990, Van Evera 1990/91; see also Kegley and Raymond 1994), it has not yet generated a sustained research program. In contrast, the first area has. The focus of this appraisal will be not so much on Waltz himself as on the neotraditional research program that has taken his proposition on balancing and investigated it empirically. This work is fairly extensive and appears to many to be both cumulative and fruitful. Specifically, the analysis will review the work of Walt (1987) and Schweller (1994) on balancing and bandwagoning, the work of Christensen and Snyder (1990) on "buck-passing" and "chain-ganging," and historical case studies that have uncovered discrepant evidence to see how these works have been treated in the field by proponents of the realist paradigm.

In addition, unlike the work on polarity, that on balancing focuses on a core area for both classical realism and neorealism. It is clearly a central proposition within the paradigm (see Vasquez 1983, 183–94), and concerns with it can be traced back to David Hume and from him to the Ancients in the West, India, and China. Given the prominence of the balance-of-power concept, a research program devoted to investigating Waltz's analysis of the balancing of power, which has attracted widespread attention and is generally well treated in the current literature, cannot fail to pass an examination of whether it is degenerating or progressive without reflecting on the paradigm as a whole—either positively or negatively.

Before beginning this appraisal it is important to keep in mind that the criterion on research programs being progressive is only one of several that can be applied to a paradigm. A full appraisal would involve the application of other criteria, such as accuracy, to all

[6] For reason of space I also do not examine formal models of the balance of power, such as those of Wagner (1986) or Niou, Ordeshook, and Rose (1989).

areas of the paradigm. Clearly, such an effort is beyond the scope of this analysis. This article provides only one appraisal, albeit a very important one, of a number that need to be conducted. As other appraisals are completed, more evidence will be acquired to make an overall assessment.

Likewise, because only the research program on balancing is examined, it can be argued that logically only conclusions about balancing (and not the other aspects of the realist paradigm) can be made. This is a legitimate position to take in that it would be illogical (as well as unfair) to generalize conclusions about one research program to others of the paradigm. Those obviously need to be evaluated separately and appraised on their own merit. They may pass or fail an appraisal based on the criterion of progressivity or on other criteria, such as empirical accuracy or falsifiability. Nevertheless, while this is true, it is just as illogical to assume in the absence of such appraisals that all is well with the other research programs.[7]

In fact, the conclusions of this study are not inconsistent with other recent work which finds fundamental deficiencies in the realist paradigm on other grounds, using different methods and addressing different questions—for example, that by Rosecrance and Stein (1993), who look at the role of domestic politics (cf. Snyder and Jervis 1993); Lebow and Risse-Kappen (1995), who examine realist and nonrealist explanations of the end of the Cold War; and George (1994), who examines the closed nature of realist thinking and its negative effects on the field.

Logically, while this analysis can only draw conclusions about the degeneracy (or progressiveness) of the research program on balancing, the implication of failing or passing this appraisal for the paradigm as a whole is not an irrelevant issue. If Waltz's neorealism is seen as reflecting well on the theoretical robustness and fertility of the realist paradigm (Hollis and Smith 1990, 66), then the failure of a research program meant to test his theory must have some negative effect on the paradigm. The question is how negative. The concluding section will return to this issue, since such matters are more fruitfully discussed in light of specific evidence rather than in the abstract.

THE BALANCING OF POWER: THE GREAT NEW LAW THAT TURNED OUT NOT TO BE SO

One of Waltz's (1979) main purposes was to explain what in his view is a fundamental law of international politics: the balancing of power. Waltz (pp. 5, 6, 9) defines theory as statements that explain laws (i.e., regularities of behavior). For Waltz (p. 117), "whenever agents and agencies are coupled by force and competition rather than authority and law," they exhibit "certain repeated and enduring patterns." These he says have been identified by the tradition of *Realpolitik*. Of these the most central pattern is balance of power, of which he says: "If there is any distinctively political theory of international politics, balance-of-power theory is it" (p. 117). He maintains that a self-help system "stimulates states to behave in ways that tend toward the creation of balances of power" (p. 118) and that "these balances tend to form whether some or all states consciously aim to establish [them]" (p. 119). This law or regularity is what the first six of the nine chapters in *Theory of International Politics* are trying to explain (see, in particular, Waltz 1979, 116–28).

The main problem, of course, is that many scholars, including many realists, such as Morgenthau ([1948] 1978, chapter 14), do not see balancing as the given law Waltz takes it to be. In many ways, raising it to the status of a law dismisses all the extensive criticism that has been made of the concept (Claude 1962; Haas 1953; Morgenthau [1948] 1978, chapter 14) (see Waltz 1979, 50–9, 117, for a review). Likewise, it also sidesteps a great deal of the theoretical and empirical work suggesting that the balance of power, specifically, is not associated with the preservation of peace (Organski 1958; Singer, Bremer, and Stuckey 1972; see also the more recent Bueno de Mesquita 1981; the earlier work is discussed in Waltz 1979, 14–5, 119).

Waltz (1979) avoided contradicting this research by arguing, like Gulick (1955), that a balance of power does not always preserve the peace because it often requires wars to be fought to maintain the balance. What Waltz does here is separate two possible functions of the balance of power—protection of the state in terms of its survival versus the avoidance of war or maintenance of the peace. Waltz does not see the latter as a legitimate prediction of balance-of-power theory. All he requires is that states attempt to balance, not that balancing prevents war.

From the perspective of Kuhn ([1962] 1970, 24, 33–4) one can see Waltz (1979) as articulating a part of the dominant realist paradigm. Waltz is elaborating one of the problems (puzzles as Kuhn [1962] 1970, 36–7, would call them) that Morgenthau left unresolved in *Politics among Nations*; namely, how and why the balance of power can be expected to work and how major a role this concept should play within the paradigm. Waltz's (1979) book can be seen as a theoryshift that places the balance of power in much more positive light than does Morgenthau (cf. 1978, chapter 14). This theoryshift tries to resolve the question of whether the balance is associated with peace by saying that it is not. Waltz, unlike Morgenthau, sees the balance as automatic; it is not the product of a particular leadership's diplomacy but of system structure. The focus on system structure and the identification of "anarchy" are two of the original contributions of Waltz (1979). These can be seen as the introduction of new concepts that bring novel facts into the paradigm. Such a shift appears progressive, but whether it proves to be so turns on whether the predictions made by the explanation can pass empirical testing.

It should come as no surprise, therefore, that the proposition on balancing is the focus of much of the research of younger political scientists influenced by

[7] I am currently engaged in a project to appraise various aspects of the realist paradigm on a variety of criteria; see Vasquez n.d.

Waltz. Walt, Schweller, Christensen and Snyder, and the historian Schroeder all cite Waltz and consciously address his theoretical proposition on balancing. They also cite and build upon the work of one another; that is, those who discuss bandwagoning cite Walt (e.g., Levy and Barrett 1991, Schweller 1994; those who talk about buckpassing cite Christensen and Snyder, 1990). More fundamentally, they generally are interested (with the exception of Schroeder, who is a critic) in working within the realist paradigm and/or defending it. They differ in terms of how they defend realism. Because they all share certain concepts, are concerned with balancing, and share a view of the world and the general purpose of trying to work within and defend the paradigm, they all can be seen as working on the same general research program. Thus, what they have found and how they have tried to account for their findings provide a good case for appraising the extent to which this particular research program is progressive or degenerating.

Balancing versus Bandwagoning

A passing comment Waltz (1979, 126) makes about his theory is that in anarchic systems (unlike domestic systems), balancing not bandwagoning (a term for which he thanks Stephen Van Evera) is the typical behavior.[8] This is one of the few unambiguous empirical predictions in his theory; Waltz (p. 121) states: "Balance-of-power politics prevail wherever two, and only two, requirements are met: that the order be anarchic and that it be populated by units wishing to survive."

The first major test is conducted by Walt (1987), who looks primarily at the Middle East from 1955 to 1979. He maintains that "balancing is more common than bandwagoning" (Walt 1987, 33). Consistent with Waltz, he argues that, in general, states should not be expected to bandwagon except under certain identifiable conditions (p. 28). Contrary to Waltz, however, he finds that they do not balance power! Instead, he shows that they balance against threat (chapter 5), while recognizing that for many realists, states should balance against power (pp. 18–9, 22–3).[9] He then extends his analysis to East-West relations and shows that if states were really concerned with power, then they would not have allied so extensively with the United States, which had a very overwhelming coalition against the USSR and its allies. Such a coalition was a result not of the power of the USSR but of its perceived threat (pp. 273–81).

Here is a clear falsification of Waltz (in the naive falsification sense of Popper 1959; see Lakatos 1970, 116), but how does Walt deal with this counterevidence or counterinstance, as Lakatos would term it? He takes a very incrementalist position. He explicitly maintains that balance of threat "should be viewed as a refinement of traditional balance of power theory" (Walt 1987, 263). Yet, in what way is this a "refinement" and not an unexpected anomalous finding, given Waltz's prediction? For Morgenthau and Waltz, the greatest source of threat to a state comes from the possible power advantages another state may have over it. In a world that is assumed to be a struggle for power and a self-help system, a state *capable* of making a threat must be guarded against because no one can be assured when it may actualize that potential. Hence, states must balance against power regardless of immediate threat. If, however, power and threat are independent, as they are perceived to be by the states in Walt's sample, then something may be awry in the realist world. The only thing that reduces the anomalous nature of the finding is that it has not been shown to hold for the central system of major states, that is, modern Europe. If it could be demonstrated that the European states balanced threat and not power, then that would be a serious if not devastating blow for neorealism and the paradigm.[10]

As it stands, despite the rhetorical veneer, Walt's findings are consistent with the thrust of other empirical research: The balance of power does not seem to work or produce the patterns that many theorists have expected it to produce. For Walt, it turns out that states balance but not for reasons of power, a rather curious finding for Waltz, but one entirely predictable given the results of previous research that found the balance of power was not significantly related to war and peace (Bueno de Mesquita 1981; see also Vasquez 1983, 183–94).

The degenerating tendency of the research program in this area can be seen in how Walt conceptualizes his findings and in how the field "refines" them further. "Balance of threat" is a felicitous phrase. The very phraseology makes states' behavior appear much more consistent with the larger paradigm than it actually is. It rhetorically captures all the connotations and emotive force of balance of power while changing it only incrementally. It appears as a refinement—insightful and supportive of the paradigm. In doing so, it strips away the anomalous nature and devastating potential of the findings for Waltz's explanation.

This problemshift, however, exhibits all four of the characteristics outlined earlier as indicative of degenerative tendencies within a research program. First, the new concept, "balance of threat," is introduced to explain why states do not balance in the way Waltz theorizes. The balance of threat concept does not appear in Waltz (1979) or in the literature before Walt introduced it in conjunction with his findings. Second, the concept does not point to any novel facts other than

[8] For Waltz (1979, 126), bandwagoning is allying with the strongest power, that is, the one capable of establishing hegemony. He maintains that such an alignment will be dangerous to the survival of states. Walt (1987, 17, 21–2) defines the term similarly but introduces the notion of threat: "*Balancing* is defined as allying with others against the prevailing threat; *bandwagoning* refers to alignment with the source of danger" (italics in original).

[9] Walt (1987, 172) concludes: "The main point should be obvious: balance of threat theory is superior to balance of power theory. Examining the impact of several related but distinct sources of threat can provide a more persuasive account of alliance formation than can focusing solely on the distribution of aggregate capabilities."

[10] Schroeder (1994a and b) provides this devastating evidence on Europe (see also Schweller 1994, 89–92).

the discrepant evidence. Third, therefore this new variant of realism does not have any excess empirical content compared to the original theory, except that it now takes the discrepant evidence and says it supports a new variant of realism.

These three degenerating characteristics open up the possibility that, when both the original balance of power proposition and the new balance of threat proposition (T and T', respectively) are taken as two versions of realism, either behavior can be seen as evidence supporting realist theory (in some form) and hence the realist paradigm or approach in general. Waltz (1979, 121) allows a clear test, because bandwagoning is taken to be the opposite of balancing. Now, Walt splits the concept of balancing into two components, either one of which will support the realist paradigm (because the second is but "a refinement" of balance-of-power theory). From outside the realist paradigm, this appears as a move to dismiss discrepant evidence and explain it away by an ad hoc theoryshift. Such a move is also a degenerating shift on the basis of the fourth indicator, because it reduces the probability that the corpus of realist propositions can be falsified. Before Walt wrote, the set of empirical behavior in which states *could* engage that would be seen as evidence falsifying Waltz's balancing proposition was much broader than it was after Walt wrote.

The danger posed by such theoryshifts can be seen by conducting a mental experiment. Would the following theoretical emendation be regarded as a progressive shift? Let us suppose that the concept of bandwagoning now becomes the focus of empirical research in its own right. Waltz (1979, 126) firmly states: "Balancing not bandwagoning is the behavior induced by the system." (Walt 1987, 32, agrees.) If someone finds bandwagoning to be more frequent, should such a finding be seen as an anomaly for Waltz's T, Walt's T', and the realist paradigm, or simply as the foundation to erect yet another version of realism (T'')? If the latter were to occur, it would demonstrate yet further degeneration of the paradigm's research program and an unwillingness of these researchers to see anything as anomalous for the paradigm as a whole.

By raising the salience of the bandwagoning concept and giving an explanation of it, Walt leaves the door open to the possibility that situations similar to the experiment may occur within the research program. Through this door walks Schweller (1994), who argues in contradiction to Walt that bandwagoning is more common than balancing. From this he weaves "an alternative theory of alliances" that he labels "balance of interests," another felicitous phrase, made even more picturesque by his habit of referring to states as jackals, wolves, lambs, and lions. Schweller (1994, 86) argues that his theory is even more realist than Waltz's, because he bases his analysis on the assumption of the classical realists—states strive for greater power and expansion—and not on security, as Waltz (1979, 126) assumes. Waltz is misled, according to Schweller (1994, 85–8), because of his status-quo bias. If he were to look at things from the perspective of a revisionist state, he would see why they bandwagon: to gain rewards (and presumably power).

Schweller (1994, 89–92), in a cursory review of European history, questions the extent to which states have balanced and argues instead that they mostly bandwagon. To establish this claim, he redefines bandwagoning more broadly than Walt; it is no longer the opposite of balancing (i.e., siding with the actor who poses the greatest threat or has the most power) but simply any attempt to side with the stronger, especially for opportunistic gain. Because the stronger state often does not pose a direct threat to every weak state, this kind of behavior is much more common and distinct from what Walt meant.

Two things about Schweller (1994) are important for the appraisal of this research program. First, despite the vehemence of his attack on the balancing proposition, this is nowhere seen as a deficiency of the realist paradigm; rather, it is Waltz's distortion of classical realism (however, see Morgenthau [1948] 1978, 194). The latter is technically true, in that Waltz raises the idea of balancing to the status of a law, but one would think that the absence of balancing in world politics, especially in European history, would have some negative effect on the realist view of the world. Certainly, Schweller's "finding" that bandwagoning is more prevalent than balancing is something classical realists, such as Morgenthau ([1948] 1978), Dehio (1961), or Kissinger (1994, 20–1, 67–8, 166–7) would find very disturbing. They would not expect this to be the typical behavior of states, and if it did occur, they would see it as a failure to follow a rational foreign policy and/or to pursue a prudent realist course (see Morgenthau [1948] 1978, 7–8).

Second, and more important, Schweller's theoryshift (T'') has made bandwagoning a "confirming" piece of evidence for the realist paradigm. So, if he turns out to be correct, his theory, which he says is even more realist than Waltz's, will be confirmed. If he is incorrect, then Waltz's version of realism will be confirmed. Under what circumstances will the realist paradigm be considered as having failed to pass an empirical test? The field is now in a position (in this research program) where any one of the following can be taken as evidence supporting the realist paradigm: balancing of power, balancing of threat, and bandwagoning. At the same time, the paradigm as a whole has failed to specify what evidence will be accepted as falsifying it—a clear violation of Popper's (1959) principle of falsifiability. Findings revealing the absence of balancing of power and the presence of balancing of threat or bandwagoning are taken by these researchers as supporting the realist paradigm; instead, from the perspective of those outside the paradigm, these outcomes should be taken as anomalies. All their new concepts do is try to hide the anomaly through semantic labeling (see Lakatos 1970, 117, 119). Each emendation tries to salvage something but does so by moving farther and farther away from the original concept. Thus, Waltz moves from the idea of a balance of power to simply balancing power, even if it does not prevent war. Walt finds that states do not balance power but oppose threats to themselves. Schweller argues that states do not balance against the stronger but more frequently bandwagon with it to take advantage of opportunities to gain rewards.

Walt and Schweller recognize discrepant evidence and explain it away by using a balance phraseology that hides the fact the observed behavior is fundamentally different from that expected by the original theory. The field hardly needs realism to tell it that states will oppose threats to themselves (if they can) or that revisionist states will seize opportunities to gain rewards (especially if the risks are low). In addition, these new concepts do not point to any novel theoretical facts; they are not used to describe or predict any pattern or behavior other than the discrepant patterns that undercut the original theory.

Ultimately, under the fourth indicator, such theory-shifts are also degenerating because they increase the probability that the realist paradigm will pass some test, since three kinds of behavior now can be seen as confirmatory. While any one version of realism (balance of power, balancing power, balance of threats, balance of interests) may be falsified, the paradigm itself will live on and, indeed, be seen as theoretically robust. In fact, the protean character of realism prevents the paradigm from being falsified because as soon as one theoretical variant is discarded, another variant pops up to replace it as the "true realism" or the "new realism."

The point is not that Walt or others are engaged in "bad" scholarship or have made mistakes; indeed, just the opposite is the case: They are practicing the discipline the way the dominant paradigm leads them to practice it. They are theoretically articulating the paradigm in a normal science fashion, solving puzzles, engaging the historical record, and coming up with new insights—all derived from neorealism's exemplar and the paradigm from which it is derived. In doing so, however, these individual decisions reflect a collective degeneration.

Even as it is, other research on bandwagoning (narrowly defined) has opened up further anomalies for the realist paradigm by suggesting that a main reason for bandwagoning (and indeed for making alliances in general) may not be the structure of the international system but domestic political considerations. Larson (1991, 86–7) argues antithetically to realism that states in a similar position in the international system and with similar relative capabilities behave differently with regard to bandwagoning; therefore, there must be some intervening variable to explain the difference. On the basis of a comparison of cases, she argues that some elites bandwagon to preserve their domestic rule (see also Strauss 1991, 245, who sees domestic considerations and cultural conceptions of world politics as critical intervening variables). Similarly, Levy and Barnett (1991, 1992) present evidence on Egypt and Third World states showing that internal needs and domestic political concerns are often more important in alliance making than are external threats. This research suggests that realist assumptions—the primacy of the international struggle for power and the unitary rational nature of the state will lead elites to formulate foreign policy strictly in accord with the national interest defined in terms of power are flawed. Theories need to take greater cognizance of the role domestic concerns play in shaping foreign policy objectives. To the extent bandwagoning is a "novel" fact (even if not a predominant pattern), it points us away from the dominant paradigm, not back to its classical formulation.

Buck-passing and Chain-ganging

The bandwagoning research program is not the only way in which the protean character of realism has been revealed. Another and perhaps even more powerful example is the way in which Christensen and Snyder (1990) have dealt with the failure of states to balance. They begin by criticizing Waltz for being too parsimonious and making indeterminate predictions about balancing under multipolarity. They then seek to correct this defect within realism, by specifying that states will engage in chain-ganging or buck-passing depending on the perceived balance between offense and defense. Chain-ganging occurs when states, especially strong states, commit "themselves unconditionally to reckless allies whose survival is seen to be indispensable to the maintenance of the balance"; buck-passing is a failure to balance and reliance on "third parties to bear the costs of stopping a rising hegemon" (Christensen and Snyder 1990, 138). The alliance pattern that led to World War I is given as an example of chain-ganging, and Europe in the 1930s is given as an example of buck-passing. The propositions are applied only to multipolarity; in bipolarity, balancing is seen as unproblematic.

This article is another example of how the realist paradigm (since Waltz) has been articulated in a normal science fashion. The authors find a gap in Waltz's explanation and try to correct it by bringing in a variable from Jervis (1978; see also Van Evera 1984). This gives the impression of cumulation and progress through further specification, especially since they have come up with a fancy title for labeling what Waltz identified as possible sources of instability in multipolarity.

A closer inspection reveals the degenerating character of their emendation. The argument that states will either engage in buck-passing or chain-ganging under multipolarity is an admission that in important instances, such as the 1930s, states fail to balance the way Waltz (1979) says they must because of the system's structure. Recall Waltz's (1979, 121) clear prediction that "balance-of-power politics will prevail wherever two, and only two, requirements are met: anarchy and units wishing to survive." Surely, these requirements were met in the period before World War II, and therefore failure to balance should be taken as falsifying evidence.

Christensen and Snyder (1990) seem to want to explain away the 1930s, in which they argue there was a great deal of buck-passing. Waltz (1979, 164–5, 167), however, never says that states will not conform overall) to the law of balancing in multipolarity, only that there are more "difficulties" in doing so. If Christensen and Snyder see the 1930s as a failure to balance properly, then this is an anomaly that needs to be explained away. The buck-passing/chain-ganging concept does that in a rhetorical flourish that grabs attention and seems persuasive. Yet, it "rescues" the theory not simply from indeterminate predictions, as Christensen and Snyder (1990, 146) put it, but explains away a critical case that the theory should have predicted.

This seems to be especially important because, contrary to what Waltz and Christensen and Snyder postulate, balancing through alliances should be more feasible under multipolarity than bipolarity, because under the latter there simply are not any other major states with whom to align. Thus, Waltz (1979, 168) says that under bipolarity *internal* balancing is more predominant and precise than external balancing. If under bipolarity there is, according to Waltz, a tendency to balance (internally, i.e., through military buildups), and under multipolarity there is, according to Christensen and Snyder, a tendency to pass the buck or chain-gang, then when exactly do we get the kind of alliance balancing that we attribute to the traditional balance of power Waltz has decreed as a law? Christensen and Snyder's analysis appears as a "protean-shift" in realism that permits the paradigm to be confirmed if states balance (internally or externally), chain-gang, or buck-pass (as well as bandwagon, see Schweller 1994). This is degenerative under the fourth indicator because the probability of falsification decreases to a very low level. It seems to increase greatly the probability that empirical tests will be passed by some form of realism.[11]

Imprecise measurement leaving open the possibility for ad hoc interpretation is also a problem with identifying buck-passing and chain-ganging. Were Britain, France, and the USSR passing the buck in the late 1930s, or were they just slow to balance? Or were Britain and France pursuing an entirely different strategy, appeasement, because of the lessons they derived from World War I? If the latter, which seems more plausible, then buck-passing is not involved at all, and the factor explaining alliance behavior is not multipolarity but an entirely different variable (see Rosecrance and Steiner 1993). What is even more troubling is that while Christensen and Snyder (1990) see pre-1939 as buck-passing and pre-1914 as chain-ganging, it seems that Britain was much more hesitant to enter the war in 1914 than in 1939, contrary to what one would expect given the logic of Christensen and Snyder's historical analysis.[12] After Hitler took Prague in March 1939, domestic public and elite opinion moved toward a commitment to war (Rosecrance and Steiner 1993, 140), but in 1914 that commitment never came before the outbreak of hostilities (see Levy 1990/91). The cabinet was split, and only the violation of Belgium tipped the balance. Thus, the introduction of the new refinement is far from a clear or unproblematic solution to the anomaly on its own terms.

The refinements of Waltz produced by the literature on bandwagoning and buck-passing are degenerating because they hide, rather than deal directly with, the seriousness of the anomalies they are trying to handle. A theory whose main purpose is to explain balancing cannot stand if balancing is not the law it says it is. Such an anomaly also reflects negatively on the paradigm as a whole. Even though Morgenthau ([1948] 1978, chapter 14) did not think the balance of power was very workable, power variables are part of the central core of his work, and he does say that the balance of power is "a natural and inevitable outgrowth of the struggle of power" and "a protective device of an alliance of nations, anxious for their independence, against another nation's designs for world domination" (Morgenthau [1948] 1978, 194, and see 173, 195–6). Waltz's (1979) theory, which has been characterized as a systematization of classical realism (Keohane 1986, 15) and widely seen as such, cannot fail on one of its few concrete predictions without reflecting badly (in some sense) on the larger paradigm in which it is embedded.

Historical Case Studies

Unlike the explicitly sympathetic work cited above, several historical case studies that focus on the balancing hypothesis give rise to more severe criticism of realist theory. Rosecrance and Stein (1993, 7) see the balancing proposition as the key prediction of structural realism. In a series of case studies, they challenge the idea that balancing power actually occurs or explains very much of the grand strategy of the twentieth-century major states they examine; to explain grand strategy for them requires examining domestic politics (Rosecrance and Stein 1993, 10, 17–21). In contradiction to structural realism, they find that balance-of-power concerns do not take "precedence over domestic factors or restraints" (Rosecrance and Stein 1993, 17). Britain in 1938, the United States in 1940, and even the Soviet Union facing Reagan in 1985 fail to meet powerful external challenges, in part because of domestic political factors (Rosecrance and Stein 1993, 18, and see the related case studies in chapters 5–7). States sometimes under- or overbalance. As Rosecrance (1995, 145) maintains, states rarely get it right—they either commit too much or too little, or they become so concerned with the periphery that they overlook what is happening to the core (see Kupchan 1994, Thompson and Zuk 1986). And, of course, they do this

[11] Of course, one may argue that Christensen and Snyder's (1994) proposition on offense-defense is falsifiable in principle, and that is true, but this points out another problem with their analysis; namely, Levy (1984) is unable to distinguish in specific historical periods whether offense or defense has the advantage (see Christensen and Snyder 1990, 139, 6 and 7). They, in turn, rely on the perception of offense and defense, but such a "belief" variable takes us away from realism and toward a more psychological-cognitive paradigm.

[12] Christensen and Snyder (1990, 156) recognize British buck-passing in 1914, but they say Britain was an outlier and "did not entirely pass the buck."

because they are not the unitary rational actors the realist paradigm thinks they are. Contrary to Waltz, and even Morgenthau, states engage in much more variegated behavior than the realist paradigm suggests.

This last point is demonstrated even more forcibly by the historian Paul Schroeder (1994a and b). He shows that the basic generalizations of Waltz—that anarchy leads states to balancing and to act on the basis of their power position—are not principles that tell the "real story" of what happened from 1648 to 1945. He demonstrates that states do not balance in a law-like manner but deal with threat in a variety of ways; among others, they hide, they join the stronger side, they try to "transcend" the problem, or they balance. In a brief but systematic review of the major conflicts in the modern period, he shows that in the Napoleonic wars, Crimean War, World War I, and World War II there was no real balancing of an alleged hegemonic threat—so much for the claim that this kind of balancing is a fundamental law of international politics. When states do resist, as they did with Napoleon, it is because they have been attacked and have no choice: "They resisted because France kept on attacking them" (Schroeder 1994a, 135; see also Schweller 1994, 92). A similar point also could be made about French, British, Soviet, and American resistance to Hitler and Japan.

Basically, Schroeder shows that the historical record in Europe does not conform to neorealists' theoretical expectations about balancing power. Their main generalizations are simply wrong. For instance, Schroeder does not see balancing against Napoleon, the prime instance in European history in which it should have occurred (see also Rosecrance and Lo 1996). Many states left the First Coalition against revolutionary France after 1793, when they should not have, given France's new power potential. Periodically, states bandwagoned with France, especially after victories, as in late 1799, when the Second Coalition collapsed. According to Schroeder (1994a, 120–1), hiding or bandwagoning, not balancing, was the main response to the Napoleonic hegemonic threat, the exact opposite of the assertions not only by Waltz but also by such long-time classical realists as Dehio (1961). For World War I, Schroeder (1994a, 122–3) argues that the balancing versus bidding for hegemony conceptualization simply does not make much sense of what each side was doing in trying to deal with security problems. With World War II, Schroeder (1994a, 123–4) sees a failure of Britain and France to balance and sees many states trying to hide or bandwagon.[13]

For Schroeder (1994a, 115, 116), neorealist theory is a misleading guide to inquiry:

> The more one examines Waltz's historical generalizations about the conduct of international politics throughout history with the aid of the historian's knowledge of the actual course of history, the more doubtful—in fact, strange—these generalizations become.... I cannot construct a history of the European states system from 1648 to 1945 based on the generalization that most unit actors within that system responded to crucial threats to their security and independence by resorting to self-help, as defined above. In the majority of instances this just did not happen.

All this suggests that the balancing of power was never the law Waltz thought it was. In effect, he offered an explanation of a behavioral regularity that never existed, except within the logic of the theory. As Schroeder (1994b, 147) concludes:

> [My point has been] to show how a normal, standard understanding of neo-realist theory, applied precisely to the historical era where it should fit best, gets the motives, the process, the patterns, and the broad outcomes of international history wrong ... it prescribes and predicts a determinate order for history without having adequately checked this against the historical evidence.

SHIRKING THE EVIDENCE AND PROVING THE POINT

How have scholars sympathetic to realism responded to Schroeder? They have sought to deny everything and done so precisely in the degenerating manner that Lakatos (1970, 116–9) predicted. The reaction by Elman and Elman (1995) to Schroeder in the correspondence section of *International Security* illustrates best the extent to which the last ten years of realist research have cumulated in degenerating problemshifts. Elman and Elman (1995) make three points against Schroeder (1994a). First, although his evidence may challenge Waltz's particular theory, it still leaves the larger neorealist approach unscathed. Second, Waltz recognizes balancing failures so that not every instance of these necessarily disconfirms his theory. Third, even if Schroeder's evidence on balancing poses a problem for Waltz, "only better theories can displace theories.... Thus, Waltz's theory should not be discarded until something better comes along to replace it" (Elman and Elman 1995, 192).

The first point somewhat misses the mark, since so much of neorealism is associated with Waltz. There remains mostly Gilpin (1981) and Krasner (1978). It is primarily Gilpin whom Elman and Elman have in mind when they argue that Schroeder's "omission of entire neo-realist literatures" leads him to fail to understand that "balancing is not the only strategy which is logically compatible with neo-realist assumptions of anarchy and self-help" (Elman and Elman 1995, 185, 186; see also Schweller 1992, 267, whom they cite).[14] They argue that for Gilpin (1981) and power transition theory "balancing is not considered a prevalent strategy, nor are balances predicted to occur repeatedly" (Elman and Elman 1995, 186). The problem with using Gilpin and the more quantitatively oriented power transition thesis of Organski and Kugler (1980) is that the two main pillars of neorealism predict contradic-

[13] Numerous other deviant cases are listed in Schroeder (1994a, 118–22, 126–9, 133–47).

[14] By saying that Schroeder leaves much of the neorealist approach unscathed, Elman and Elman (1995) seem to fall into the trap of assuming that Gilpin (1981) is empirically accurate unless proven otherwise. In fact, as related to security questions, Gilpin (1981) has not been extensively tested, and existing tests are not very encouraging (see Spiezio 1990, as well as Boswell and Sweat (1991) and the discussion in Vasquez 1993, 93–8).

tory things. Thus, between Waltz and Gilpin, threat can be handled by either balancing or not balancing. It certainly is not a very strong defense of neorealism to say that opposite behaviors are both logically compatible with the assumptions of anarchy.

The Elmans are technically correct that evidence against balancing does not speak against all the larger realist paradigm in that neorealism also embodies Gilpin. But it is this very correctness that proves the larger point being made here and illustrates what so worried Lakatos about degenerating research programs. At the beginning of this article, four indicators of a degenerating research program were presented. Elman and Elman (1995) serves as evidence that all these are very much in play within the field. On the basis of their defense of neorealism and the review of the literature above, it will be shown that the protean nature of realism, promulgated by the proliferation of auxiliary hypotheses to explain away discrepant evidence, has produced an unwillingness to specify what evidence would in principle lead to a rejection of the paradigm. The result has been a continual theoretical articulation but in the context of a persistent dearth of strong empirical findings.

Using Gilpin and power transition in the manner of the Elmans is degenerating because permitting the paradigm to be supported by instances of either "balancing" or "not balancing" reduces greatly the probability of finding any discrepant evidence. As if this were not enough to cover all sides of the bet, Elman and Elman (1995, 187–8) maintain that, within the neorealist assumption of self-help, threat can be handled by bandwagoning, expansion, preventive war, balancing, hiding, and even what Schroeder has labeled "transcending."[15] In other words, there is always some behavior (in dealing with threat) that will prove realism correct, even though most versions will be shown to be incorrect, and even though neorealists "often consider balancing to be the most successful strategy for most states most of the time" (Elman and Elman 1995, 187). But if this caveat is the case, then why do states not regularly engage in this behavior? Elman and Elman rightly capture the theoretical robustness of the realist paradigm—showing that Waltz, Gilpin, and others are part of the paradigm—but they fail to realize the damning protean portrayal they give of its research program and how this very theoretical development makes it difficult for the paradigm to satisfy the criterion of falsifiability.

Instead, they conclude about Schroeder's (1994a) historical evidence that "no evidence could be more compatible with a neo-realist reading of international relations" (Elman and Elman 1995, 184). They conclude this because each of these strategies (bandwagoning, etc.) does not challenge the realist conception of a rational actor behaving in a situation of competition and opportunity. For them, so long as states choose strategies that are "consistent with their position in the global power structure and pursue policies that are likely to provide them with greater benefits than costs" (Elman and Elman 1995, 184), then this is seen as evidence supporting the broad realist approach. Only Wendt's (1992) claim that states could be "other-regarding" as opposed to "self-regarding" is seen as discrepant evidence (see also Elman 1996, Appendix, Diagram 1). Basically, these are "sucker bets" of the "I win, you lose" variety. Let it be noted that these are not bets that Elman and Elman are proposing; they are merely reporting what, in effect, the entire realist research program has been doing—from Walt, to Christensen and Snyder, to Schweller, and so forth. Collectively, the realist mainstream has set up a situation that provides a very narrow empirical base on which to falsify the paradigm.

What kinds of political actors would, for example, consciously pursue policies that are "likely to provide" them with greater costs than benefits? To see only "other-regarding" behavior as falsifying leaves a rather vast and variegated stream of behaviors as supportive of the paradigm. Schroeder (1995, 194) has a legitimate complaint when he says, in reply: "The Elman argument . . . appropriates every possible tenable position in IR theory and history for the neo-realist camp." He concludes: "Their whole case that history fits the neo-realist paradigm falls to the ground because they fail to see that it is their neo-realist assumptions, as they understand and use them, which simply put all state action in the state system into a neo-realist mold and neo-realist boxes, *by definition*" (Schroeder 1995, 194, emphasis in the original).

Instead of defending the paradigm, Elman and Elman (1995) expose the degenerating nature of its research program and the field's collective shirking of the evidence through proteanshifts. Many neotraditionalists, such as Mearsheimer (1990), have eschewed quantitative evidence challenging the adequacy of the realist paradigm; if realists will now refuse to accept historical evidence as well, then what kind of evidence will they accept as falsifying their theories? Only "other-regarding" behavior? That simply will not do.

The cause of this problem is the lack of rigor in the field in appraising theories. The nature of the problem can be seen in Elman and Elman's (1995) second point against Schroeder. Drawing upon Christensen and Snyder (1990), they note that balancing under multipolarity, for Waltz, is more difficult than balancing under bipolarity: "Thus Schroeder's finding that states failed to balance prior to World War I (pp. 122–3) and World War II (pp. 123–4) does not disconfirm Waltz's argument. . . . In short, a failure to balance is not a failure of balance of power theory if systemic conditions are likely to generate this sort of outcome in the first place" (Elman and Elman 1995, 190–1). This sets up a situation in which any failure to balance under multipolarity can be taken as confirmatory evidence because, according to Elman and Elman (1995, 90), "Waltz's theory also predicts balancing *failures*" (emphasis in the original). This again poses an "I win, you lose" bet. If the periods before World War I and World War II are not legitimate tests of Waltz's prediction of

[15] Transcending is seen by Schroeder (1994a) as particularly discrepant for realism, but Elman and Elman (1995, 188) view it as part of the realist approach.

balancing, then what would be? The implication is that balancing more easily occurs under bipolarity, but here external balancing is structurally impossible by definition. If this is the case, how is balancing a "law," or the main outcome of anarchy? This is especially problematic because there is a tendency in Waltz to see only the post-1945 period as a true bipolarity (see Nye 1988, 244), which means the rest of history is multipolar and subject to balancing failures.

In the end, Elman and Elman (1995, 192) concede that Waltz does believe that, "on aggregate," states should balance, so "Schroeder's evidence that states rarely balance does indeed pose a problem for Waltz's theory." They conclude, however, by citing Lakatos—only better theories can displace theories—and therefore Waltz's theory should not be discarded until something better comes along. They then outline a general strategy for improving the theory, namely, adding variables, identifying the domain to which it is applicable, and broadening definitions (especially of threat). All these, however, are precisely the tactics that have produced the degenerating situation the field now faces. Thus, they say, by broadening the definition of threat to include internal threats from domestic rivals, decision makers could still be seen as balancing, and bandwagoning "would not necessarily disconfirm the prediction that balancing is more common" (Elman and Elman 1995, 192). This would take the discrepant evidence of Levy and Barnett (1991, 1992) and of Larson (1991) and make it confirmatory. This is precisely the kind of strategy that Lakatos (1970, 117–9) decried.

What is also evident from this appraisal of the realist paradigm is that Lakatos's (1970, 119) comment that "there is no falsification before the emergence of a better theory" can play an important role in muting the implications of a degenerating research program, especially when alternative paradigms or competing mid-range theories are ignored, as has been the case in international relations. There have been too many empirical failures and anomalies, and theoretical emendations have taken on an entirely too ad hoc nonfalsifying character for adherents to say that the paradigm cannot be displaced until there is a clearly better theory available. Such a position makes collective inertia work to the advantage of the dominant paradigm and makes the field less rather than more rigorous.

If one wants to take the very cautious position that Schroeder's historical evidence affects only Waltz, one should not then be incautious and assume that other research programs within the realist paradigm are doing fine. A more consistent position would be to hold this conclusion in abeyance until all aspects of the paradigm are appraised. The lesson from Schroeder's (1994a and b) discrepant evidence should *not* be that his "article leaves the general neo-realist paradigm unscathed" (Elman and Elman 1995, 192), but that a major proposition of the paradigm has failed to pass an important historical test.

WHERE DO WE GO FROM HERE?

It seems that the internal logic of the Lakatos rules requires that a warning flag on the degenerating direction of the research program on balancing be raised. Theorists should be aware of the pitfalls of setting up realist variants that produce a "heads, I win; tails, you lose" situation, which makes realism nonfalsifiable. In addition, greater efforts need to be made in specifying testable differences between realist and nonrealist explanations before the evidence is assessed, so as to limit the use of ex post facto argumentation that tries to explain away discrepant evidence.

If one accepts the general thrust of the analysis that the neotraditional research program on balancing has been degenerating, then the question that needs to be discussed further is the implications of this for the wider paradigm. Two obvious conclusions are possible. A narrow and more conservative conclusion would try to preserve as much of the dominant paradigm as possible in face of discrepant evidence. A broader and more radical conclusion would take failure in this one research program as consistent with the assessments of other studies and thus as an indicator of a deeper, broader problem. It is not really necessary that one conclusion rather than the other be taken by the entire field, since what is at stake here are the research bets individuals are willing to take with their own time and effort. In this light, it is only necessary to outline the implications of the two different conclusions.

The narrow conclusion is that Waltz's attempt to explain what he regards as the major behavioral regularity of international politics was premature because states simply do not engage in balancing with the kind of regularity that he assumes. It is the failure of neotraditional researchers and historians to establish clearly the empirical accuracy of Waltz's balancing proposition that so hurts his theory. If the logical connection between anarchy (as a systemic structure) and balancing is what Waltz claims it to be, and states do not engage in balancing, then this empirical anomaly must indicate some theoretical deficiency.

The neotraditional approach to date has muted the implications of the evidence by bringing to bear new concepts. The argument presented here is that such changes are primarily semantic and more clearly conform to what Lakatos calls degenerating theoryshifts than to progressive theoryshifts. If this is accepted, then at minimum one would draw the narrow conservative conclusion that the discrepant evidence (until further research demonstrates otherwise) is showing that states do not balance in the way Waltz assumes they do. Realists then can concentrate on other research programs within the paradigm without being susceptible (at least on the basis of this analysis) to the charge of engaging in a degenerating research program. Those who continue to mine realist inquiry, however, should pay more attention to the problem of degeneration in making theoretical reformulations of realism. Specifically, scholars making theoryshifts in realism should take care to ensure that these are not just proteanshifts.

The implication of the broader and more radical conclusion is to ask why a concept so long associated with realism should do so poorly and so misguide so many theorists. Could not its failure to pass neotraditional and historical "testing" (or investigation) be an indicator of the distorted view of world politics that the paradigm imposes on scholars? Such questions are reasonable to ask, especially in light of appraisals that have found other aspects of realism wanting (see Lebow and Risse-Kappen 1995, Rosecrance and Stein 1993, Vasquez 1983), but they are not the same as logically compelling conclusions that can be derived from the analysis herein. It has been shown only that one major research program, which has commanded a great deal of interest, seems to be exhibiting a degenerating tendency.

Such a demonstration is important in its own right, particularly if analysts are unaware of the collective effect of their individual decisions. In addition, it shows that what admirers of the realist paradigm have often taken as theoretical fertility and a continuing ability to provide new insights is not that at all, but a degenerating process of reformulating itself in light of discrepant evidence.

Regardless of whether a narrow or broad conclusion is accepted, this analysis has shown that the field needs much more rigor in the interparadigm debate. Only by being more rigorous both in testing the dominant paradigm and in building a new one that can explain the growing body of counterevidence as well as produce new nonobvious findings of its own will progress be made.

REFERENCES

Banks, Michael. 1985. "The Inter-Paradigm Debate." In *A Handbook of Current Theory*, ed. Margot Light and A. J. R. Groom. London: Frances Pinter.

Boswell, Terry, and Mike Sweat. 1991. "Hegemony, Long Waves, and Major Wars: A Time Series Analysis of Systemic Dynamics, 1496–1967." *International Studies Quarterly* 35(June):123–49.

Bueno de Mesquita, Bruce. 1981. "Risk, Power Distributions, and the Likelihood of War." *International Studies Quarterly* 25(December):541–68.

Claude, Inis L., Jr. 1962. *Power and International Relations*. New York: Random House.

Christensen, Thomas J., and Jack Snyder. 1990. "Chain Gangs and Passed Bucks: Predicting Alliance Patterns in Multipolarity." *International Organization* 44(Spring):137–68.

Dehio, Ludwig. 1961. *The Precarious Balance*. New York: Vintage.

Doyle, Michael W. 1986. "Liberalism and World Politics." *American Political Science Review* 80(December):1151–69.

Elman, Colin. 1996. "Horses for Courses: Why *Not* Neorealist Theories of Foreign Policy?" *Security Studies* 6(Autumn):7–53.

Elman, Colin, and Miriam Fendius Elman. 1995. "History vs. Neorealism: A Second Look." *International Security* 20(Summer):182–93.

George, Jim. 1994. *Discourses of Global Politics: A Critical (Re)Introduction to International Relations*. Boulder, CO: Lynne Rienner.

Gilpin, Robert. 1981. *War and Change in World Politics*. Cambridge: Cambridge University Press.

Gulick, Edward V. 1955. *Europe's Classical Balance of Power*. New York: Norton.

Haas, Ernst B. 1953. "The Balance of Power: Prescription, Concept, or Propaganda?" *World Politics* 5(April):442–77.

Hollis, Martin, and Steve Smith. 1990. *Explaining and Understanding International Relations* Oxford: Clarendon Press.

Jervis, Robert. 1978. "Cooperation under the Security Dilemma." *World Politics* 30(January):167–214.

Kegley, Charles W., Jr., ed. 1995. *Controversies in International Relations Theory: Realism and the Neoliberal Challenge*. New York: St. Martin's.

Kegley, Charles W., Jr., and Gregory A. Raymond. 1994. *A Multipolar Peace? Great-Power Politics in the Twenty-First Century*. New York: St. Martin's.

Keohane, Robert O. [1983] 1989. "Theory of World Politics: Structural Realism and Beyond." In *Political Science: The State of the Discipline*, ed. Ada W. Finifter. Washington, DC: American Political Science Association. Reprinted in Robert O. Keohane. *International Institutions and State Power*. Boulder, CO: Westview.

Keohane, Robert O. 1986. "Realism, Neorealism and the Study of World Politics." In *Neorealism and Its Critics*, ed. Robert O. Keohane. New York: Columbia University Press.

Kissinger, Henry. 1994. *Diplomacy*. New York: Simon & Schuster.

Krasner, Stephen D. 1978. *Defending the National Interests*. Princeton, NJ: Princeton University Press.

Kuhn, Thomas S. [1962] 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kupchan, Charles. 1994. *The Vulnerability of Empire*. Ithaca, NY: Cornell University Press.

Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes." In *Criticism and the Growth of Knowledge*, ed. Imre Lakatos and Alan Musgrave. Cambridge: Cambridge University Press.

Lapid, Yosef. 1989. "The Third Debate: On the Prospects of International Theory in a Post-Positivist Era." *International Studies Quarterly* 33(September):235–54.

Larson, Deborah Welch. 1991. "Bandwagon Images in American Foreign Policy: Myth or Reality?" In *Dominoes and Bandwagons*, ed. Robert Jervis and Jack Snyder. New York: Oxford University Press.

Lebow, Richard Ned, and Thomas Risse-Kappen, eds. 1995. *International Relations Theory and the End of the Cold War*. New York: Columbia University Press.

Levy, Jack S. 1984. "The Offensive/Defensive Balance of Military Technology: A Theoretical and Historical Analysis." *International Studies Quarterly* 28(June):219–38.

Levy, Jack S. 1990/91. "Preferences, Constraints, and Choices in July 1914." *International Security* 15(Winter):151–86.

Levy, Jack S., and Michael N. Barnett. 1991. "Domestic Sources of Alliances and Alignments: The Case of Egypt, 1962–1973." *International Organization* 45(Summer):369–95.

Levy, Jack S., and Michael N. Barnett. 1992. "Alliance Formation, Domestic Political Economy, and Third World Security." *Jerusalem Journal of International Relations* 14(December):19–40.

Masterman, Margaret. 1970. "The Nature of a Paradigm." In *Criticism and the Growth of Knowledge*, ed. Imre Lakatos and Alan Musgrave. Cambridge: Cambridge University Press.

Mearsheimer, John J. 1990. "Back to the Future: Instability in Europe after the Cold War." *International Security* 15(Summer): 5–56.

Morgenthau, Hans J. [1948] 1978. *Politics among Nations*. 5th rev. ed. New York: Knopf.

Niou, Emerson, Peter C. Ordeshook, and Gregory F. Rose. 1989. *The Balance of Power: Stability in International Systems*. New York: Cambridge University Press.

Nye, Joseph S., Jr. 1988. "Neorealism and Neoliberalism." *World Politics* 40(January):235–51.

Nye, Joseph S., Jr. 1993. *Understanding International Conflicts*. New York: Harper Collins.

Olson, William C., and A. J. R. Groom. 1991. *International Relations Then and Now: Origins and Trends in Interpretation*. London: Harper Collins Academic.

Organski, A. F. K. 1958. *World Politics*. New York: Knopf.

Organski, A. F. K., and Jacek Kugler. 1980. *The War Ledger*. Chicago: University of Chicago Press.

Popper, Karl. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.

Rosecrance, Richard. 1995. "Overextension, Vulnerability, and Conflict." *International Security* 19(Spring):145–63.

Rosecrance, Richard, and Chih-Cheng Lo. 1996. "Balancing, Stability, and War: The Mysterious Case of the Napoleonic International System." *International Studies Quarterly* 40(December):479–500.

Rosecrance, Richard, and Arthur Stein, eds. 1993. *The Domestic Bases of Grand Strategy*. Ithaca, NY: Cornell University Press.

Rosecrance, Richard, and Zara Steiner. 1993. "British Grand Strategy and the Origins of World War II." In *The Domestic Bases of Grand Strategy*, ed. Richard Rosecrance and Arthur Stein. Ithaca, NY: Cornell University Press.

Schroeder, Paul W. 1994a. "Historical Reality vs. Neo-realist Theory." *International Security* 19(Summer):108–48.

Schroeder, Paul W. 1994b. *The Transformation of European Politics, 1763–1848*. Oxford: Clarendon Press.

Schroeder, Paul W. 1995. "History vs. Neo-realism: A Second Look, The Author Replies." *International Security* 20(Summer):193–5.

Schweller, Randall L. 1992. "Domestic Structure and Preventive War: Are Democracies More Pacific?" *World Politics* 44(January): 235–69.

Schweller, Randall L. 1994. "Bandwagoning for Profit: Bringing the Revisionist State Back In." *International Security* 19(Summer):72–107.

Simowitz, Roslyn, and Barry Price. 1990. "The Expected Utility Theory of Conflict: Measuring Theoretical Progress." *American Political Science Review* 84(June):439–60.

Singer, J. David, Stuart Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820–1965." In *Peace, War and Numbers*, ed. Bruce Russett. Beverly Hills, CA: Sage.

Smith, Steve. 1995. "The Self-Images of a Discipline: A Genealogy of International Relations Theory." In *International Relations Theory Today*, ed. Ken Booth and Steve Smith. Cambridge, MA: Polity Press.

Snyder, Jack, and Robert Jervis, eds. 1993. *Coping with Complexity in the International System*. Boulder, CO: Westview.

Spiezio, K. Edward. 1990. "British Hegemony and Major Power War, 1815–1935: An Empirical Test of Gilpin's Model of Hegemonic Governance." *International Studies Quarterly* 34 (June):165–81.

Strauss, Barry S. 1991. "Of Balances, Bandwagons, and Ancient Greeks." In *Hegemonic Rivalry: From Thucydides to the Nuclear Age*, ed. Richard Ned Lebow and Barry S. Strauss. Boulder, CO: Westview.

Thompson, William R., and Gary Zuk. 1986. "World Power and the Strategic Trap of Territorial Commitments." *International Studies Quarterly* 30(September):249–67.

Van Evera, Stephen. 1984. "The Cult of the Offensive and the Origins of the First World War." *International Security* 9(Summer): 58–107.

Van Evera, Stephen. 1990/91. "Primed for Peace: Europe after the Cold War." *International Security* 15(Winter):7–57.

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

Vasquez, John A. 1992. "World Politics Theory." In *Encyclopedia of Government and Politics*, ed. Mary Hawkesworth and Maurice Kogan. London: Routledge.

Vasquez, John A. 1993. *The War Puzzle*. Cambridge: Cambrige University Press.

Vasquez, John A. 1995. "The Post-Positivist Debate: Reconstructing Scientific Enquiry and International Relations Theory after Enlightenment's Fall." In *International Relations Theory Today*, ed. Ken Booth and Steve Smith. Cambridge, MA: Polity Press.

Vasquez, John A. 1996. "When Are Power Transitions Dangerous? An Appraisal and Reformulation of Power Transition Theory." In *Parity and War*, ed. Jacek Kugler and Douglas Lemke. Ann Arbor: University of Michigan Press.

Vasquez, John A. N.d. *The Power of Power Politics: From Classical Realism to Neotraditionalism*. Cambridge: Cambridge University Press. Forthcoming.

Wagner, R. Harrison. 1986. "The Theory of Games and the Balance of Power." *World Politics* 38(July):546–76.

Walt, Stephen M. 1987. *The Origins of Alliances*. Ithaca, NY: Cornell University Press.

Waltz, Kenneth N. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley.

Wayman, Frank W., and Paul F. Diehl, eds. 1994. *Reconstructing Realpolitik*. Ann Arbor: University of Michigan Press.

Wendt, Alex. 1992. "Anarchy Is What States Make of It: The Social Construction of Power Politics." *International Organization* 46(Spring):391–425.