

## 4. Beyond Story-Telling: Designing Case Study Research in International Environmental Policy

Ronald Mitchell and Thomas Bernauer

Scholars have employed a range of methods to explore international environmental politics and policy (IEP). A small, but increasing, number have approached the subject quantitatively (Sprinz and Vaahtoranta 1994; Underdal 2001; Mitchell 2002a). Most, however, have used qualitative methods. In doing so, they have varied in whether their goals are descriptive (what did happen?), predictive (what will happen?), normative (what should happen?), or explanatory (what made this happen?). Like the rest of the international relations community, IEP scholars also vary in theoretical proclivities, ranging along the constructivist-rationalist continuum (Ruggie 1998). Scholars' different goals and proclivities have also led to debates about what qualitative research can, and should try to, accomplish. These debates are unlikely to be resolved soon, nor do we aspire to do so here. We believe that IEP scholars are more likely to understand the complexities of IEP accurately and fully if they use, as a community if not as individuals, a range of methods.

Our goal is to aid those IEP students and scholars who, at least some of the time, use qualitative methods to pursue explanatory goals within a rationalist paradigm. Descriptive analyses of existing international environmental issues, processes, negotiations, or agreements and prescriptive analyses of alternative policies clearly play important political and intellectual roles in IEP. Our argument is limited to noting that they usually contribute little to our understanding of causal relationships. For ease of expo-

sition and to clarify that our argument is not intended to extend to all qualitative scholarship, we use QER (qualitative, explanatory, rationalist) to refer to the explanatory and rationalist subset of qualitative scholars and approaches. For these scholars, we seek to identify general principles to emulate and pitfalls to avoid in qualitative studies aimed at drawing causal inferences regarding independent and dependent variables from one or a small set of IEP cases that can be generalized to a larger class of cases.

After a brief look at the opportunities and challenges of case study research in IEP we outline six steps that promote the drawing of systematic, rigorous, and theoretically and empirically informative conclusions from qualitative research in this area. We introduce the reader to some key problems and possible solutions in qualitative case study research. However, the focus of this chapter is primarily on empirical applications of this method in IEP. Qualitative case study research is treated in more generic terms by Bennett (chap. 2, this vol.). We are motivated by our view that case studies of IEP too frequently make methodological errors that lead to unnecessarily weak or inaccurate conclusions. Rather than highlight the many available examples of such errors, we seek to identify positive empirical or hypothetical examples of how researchers have or could have avoided such errors.

### **Opportunities and Challenges of Causal Research in IEP**

Causal analysis is central to much IEP research, if only because most scholars working in the field bring to their research a normative commitment to reducing human degradation of the environment. To actualize that commitment through research that links knowledge to action requires the researcher to identify causes of environmentally harmful outcomes with sufficient accuracy that governmental or private decision makers using their findings can be confident regarding which actions will avert such outcomes and under which conditions. It involves carefully identifying the political, economic, and social forces that cause international environmental problems; why some environmental problems get internationalized while others do not; and why solutions are devised for some international problems but not for others. It also engages questions of why some intergovernmental, corporate, and nongovernmental policies and behaviors mitigate these problems while others do not and what determines global

society's success at evaluating and improving its attempts to protect the global environment (Mitchell 2002b).

Although similar questions inform research on security, international political economy, and human rights, the character of international environmental policy presents somewhat different opportunities and challenges to qualitative research. Interstate governance is less centralized than in any of these other arenas—several hundred multilateral agreements and hundreds more bilateral agreements have been established, most involving their own secretariats and with surprisingly little coordination across treaties or regimes (Mitchell *in press*). In regime formation and effectiveness research, the availability of numerous cases has facilitated edited volumes that evaluate multiple regimes (Sand 1992; Haas, Keohane, and Levy 1993; Keohane and Levy 1996; Brown Weiss and Jacobson 1998; Victor, Raustiala, and Skolnikoff 1998; Young 1998, 1999; Miles et al. 2001). The best of these and related works produce relatively rigorous comparative case studies, but many could be described more accurately as parallel analysis. In part this reflects the nature of multiauthor edited volumes, but it also reflects a deeper belief of many IEP researchers that differences in the nature of the underlying environmental problems preclude meaningful comparison across environmental problems or the social efforts to mitigate them. Recent work has begun to engage this assumption, taking up the task of defining metrics, identifying cases, and developing data in ways that would allow meaningful comparisons and inclusion in a common database for quantitative study (Breitmeier et al. 1996; Helm and Sprinz 2000; Miles et al. 2001; Mitchell 2002a; Young 2001).

International environmental politics also poses different challenges than other issues in the lack of consensus regarding what constitutes policy effectiveness, a dependent variable for many researchers working in the field (Mitchell 2002b). As in other arenas, some view negotiation of an international agreement as evidence of meaningful cooperation, while others believe compliance is necessary. Unlike most other areas, however, there has been ongoing debate in the subfield regarding whether even compliance is an appropriate object of study, since compliance is neither a necessary nor sufficient condition to achieve what many scholars consider to be the ultimate goal of international environmental cooperation, namely, significant environmental improvement. We illustrate and attempt to address these and related challenges in the rest of this chapter.

## Performance Criteria and Research Steps

The literature on designing and carrying out systematic, rigorous QER research (George and Bennett forthcoming; Tetlock and Belkin 1996; Yin 1994, 54–101; King, Keohane, and Verba 1994; Ragin 1994) points to five criteria (phrased in table 1 as questions) that characterize innovative contributions to our understanding of the sources of variation in IEP. Meeting these criteria necessitates attention to the six tasks discussed later. Although delineated in logical order, research rarely does, and often should not, proceed in linear fashion; high-quality research often requires an iterative path through these steps. But those conducting QER research should address each of these tasks explicitly to produce findings that compel, convince, and contribute.<sup>1</sup>

### Identifying an Important Theoretical Question

Innovative QER research on IEP should address existing theoretical debates in the field while attending to current policy concerns (King, Keohane, and Verba 1994, 15). Research efforts should target unresolved debates, untested theoretical claims including those about the merits of unassessed policy “innovations,” or previously uninvestigated relation-

**TABLE 1. Criteria for High Quality QER Research**

Criteria	Question
Construct validity	Does the collected empirical information accurately capture the concepts or variables contained in the theoretical model or propositions nominally being investigated?
Internal validity	Does the analytic method demonstrate that, for each hypothesized causal relationship, variation observed in the independent variable correlates with observed variation in the dependent variable, and that no other variables provide a more plausible explanation of variation in the dependent variable?
External validity	Has the researcher accurately identified the boundary between the class of cases to which the findings can be validly generalized and beyond which valid generalizations are unlikely?
Reliability	Could other researchers replicate the research techniques used, e.g., data collection and analytic methods, and, having done so, arrive at the same results?
Progressive research	Does the research contribute to a larger, cumulative research program?

ships. Appropriate targeting of research questions requires a sophisticated understanding of existing theory and existing empirical patterns.

Familiarity with extant theoretical literatures helps the researcher identify whether variation in a particular variable is considered of interest to other scholars, identifies the factors that explain that variation or are explained by it, and allows a researcher to frame research questions in ways that help build a progressive research program while engaging appropriate scholarly communities (Lakatos 1970). Early scholarship in IEP often assumed issues were undertheorized because no one had explicitly sought to explain the phenomena of interest. Research on IEP has come of age over the last decade, however. Extensive theoretical work already exists regarding the roles of nongovernmental actors and movements in IEP (Lipschutz and Conca 1993; Princen and Finger 1994; Wapner 1996; Dawson 1996; Clapp 1998), the domestic sources of international environmental policy (Schreurs and Economy 1997; O'Neill 2000; DeSombre 2000), international regime formation (Susskind, Siskind, and Breslin 1990; Lipschutz 1991; Sjostedt 1993; Young and Osherenko 1993; Meyer et al. 1997; Young 1998), and international regime effects (Sand 1992; Haas, Keohane, and Levy 1993; Keohane and Levy 1996; Brown Weiss and Jacobson 1998; Victor, Raustiala, and Skolnikoff 1998; Young 1998, 1999; Miles et al. 2001).

At this point, there is no dearth of theories susceptible to evaluation with QER methods. Even scholars interested in exploring previously uninvestigated issues in IEP can begin by identifying how general theories of international relations—for example, realism, institutionalism, critical theory, constructivism—have answered the central research issue, or would answer it given the core principles of their theoretical arguments. Thus, early scholarship on environmental regime compliance derived predictions from broader theories of international relations rather than “starting new” (Young 1989, chap. 3; Haas 1990, chap. 2; Mitchell 1994, chap. 2). Most scholars now start from the assumption that general theories of international relations apply as well to IEP as to other policy domains, and use evidence that they do not to support such conclusions as military or economic power and interests explain less variation in IEP outcomes than in security or that prevailing theories are less generalizable than claimed, rather than that no relevant theory exists.

Empirical puzzles provide a useful way to frame QER research. Highlighting a contradiction between outcomes predicted by theory and those

actually observed engages the reader and grounds theory. Puzzles can also stem from competing theoretical explanations for the same outcome, from untested theoretical predictions, or from the absence of any theories purporting to explain an observed outcome. Puzzles that use a particular outcome as a way of generating a debate between competing theories are particularly valuable since their findings are assured of refuting one theory and lending support to another or refuting both, regardless of the particular empirical findings. Testing one theory's prediction without a competing theory's prediction is riskier: the researcher cannot know beforehand whether cases selected will disconfirm a theory and so may end up with results that fail to falsify a theory yet, by definition, cannot prove it and, because of the small number of observations in QER research, cannot even provide strong support for it (Popper 1968).

Defining the research in terms that view the case(s) studied as instances in a larger class of cases (see Bennett, chap. 2, this vol.) is also crucial to QER analysis of IEP.<sup>2</sup> Thus, Princen's (1996) argument that the ivory trade ban established a norm against purchasing ivory was of interest not simply to those concerned about rhinos and elephants because he framed it as an example of how bans establish ecological norms more effectively than other policies. Several edited volumes (Young and Osherenko 1993; Haas, Keohane, and Levy 1993; Keohane and Levy 1996; Brown Weiss and Jacobson 1998; Victor, Raustiala, and Skolnikoff 1998) also show how case studies contribute to cumulative knowledge when cases with substantive appeal are analyzed as examples of larger classes of cases. In short, successful QER studies in IEP must match a generalizable theoretical question with a corresponding empirical one (Stinchcombe 1968).

Given such a theoretical foundation, QER research in IEP usually fits into one of three possible modes. Researchers can attempt to explain change in a specified dependent variable (DV), being initially agnostic about whether any particular factor was its major cause. Thus, the Social Learning Group set out to understand why "social learning" (defined as improvement in a society's ability to manage environmental risks) occurred in specific countries with respect to certain environmental problems at particular points in time but not in other countries, with respect to other problems, or at other times. They attributed the observed variation to a complex interplay of ideas, interests, and institutions, rather than any one factor (2001a, 2001b). Researchers may also specify both independent variables (IVs) and dependent variables at the outset. Much of the work to

date on regime compliance and effectiveness has taken this shape, finding that international environmental regimes influence environmental policies and behaviors only under specified conditions (Brown Weiss and Jacobson 1998; Victor, Raustiala, and Skolnikoff 1998; Miles et al. 2001). A collection of QER studies of international environmental aid found that it had its intended effect only infrequently (Keohane and Levy 1996). Current research on global environmental science assessments demonstrates that their influence on international politics depends on their being simultaneously salient, legitimate, and credible to multiple audiences (Mitchell et al. forthcoming). Finally, researchers can analyze the effects of a specific IV, as evident in much of Oran Young's recent work that seeks to examine the full range of effects of international environmental regimes, examining direct and indirect effects, effects that are internal to the problem being addressed and external to it, and effects that are intended and positive as well as those that are unintended and negative, an approach also followed by those interested in the pathologies of international organizations (Young 1999, 15; Barnett and Finnemore 1999).

### Developing Hypotheses and Identifying Variables

Efforts to draw causal inferences in IEP requires that hypotheses "can be shown to be wrong as easily and quickly as possible" and that criteria for what constitutes falsification be identified before research begins (King, Keohane, and Verba 1994, 100). In developing hypotheses, the researcher must carefully identify independent, control, and dependent variables, their potential values, and their theorized causal relationships.<sup>3</sup> Consider an effort to apply QER methods to the claim that "a country's culture influences how likely it will be to join an environmental regime." The DV (regime membership) and IV (culture) are clear, as are the potential values of the DV (ranging from extremely likely to extremely unlikely), but the values of the IV are not. To make this hypothesis falsifiable, the researcher must categorize "culture" into at least two values that she predicts correspond to higher and lower likelihoods of membership. Whether categorized as indigenous and nonindigenous; strongly environmental, weakly environmental, and nonenvironmental; or by some other taxonomy, evaluating this hypothesis requires defining categories for culture.

Most recent work on regime compliance and effectiveness carefully specifies both the predicted influence of each IV on the DV and potential

causal pathways or mechanisms by which these variables may be operating (Brown Weiss and Jacobson 1998; Young and Osherenko 1993; Young 1999; Wettestad 1999; Miles and Underdal 2000). Extant theory is crucial in developing hypotheses and designing a convincing study, since it identifies which IVs need to be evaluated (and which do not) to explain variation in a DV convincingly. Familiarity with previous research, both theoretical and empirical, clarifies which variables should be included as controls or evaluated as alternative explanations of the DV, and it also clarifies effective and convincing ways to observe or operationalize all the variables in the study. In short, familiarity with earlier work identifies which variables to include and how to include them, as well as which variables can be ignored as not relevant to the study at hand.

Much QER research in IEP seeks to identify the one or more independent variables that caused the observed variation of a dependent variable in a particular case from the plethora of other potentially explanatory variables. Identifying such causal inferences requires that the researcher evaluate more observations than independent variables. Recalling Bennett's distinction between cases and observations (chap. 2, this vol.), note that it does not require multiple cases. Many single case studies draw well-supported inferences by making multiple observations of the IVs and DV in their case over time. Others have developed compelling findings by carefully constructing counterfactuals that are sufficiently plausible to serve as additional "observations" (Biersteker 1993; Fearon 1991). The key point here is that to identify the influence of a single IV on a DV requires at least two observations—we believe a minimal standard for claiming that A caused B involves demonstrating that A and B were observed together but also that not-A and not-B were observed together. Many case studies in IEP and elsewhere that are unconvincing because they demonstrate only that A and B were observed together could be significantly improved by the simple process of adding evidence that prior to A's coming to pass, B was not observed.

Indeed, case studies of single nongovernmental organizations (NGOs), single negotiations, or single regimes become quite convincing when they explicitly identify the several, sometimes numerous, potential observations within that case. Young and his colleagues call these observations "snapshots," examining both regime structure (the IV) and behavior patterns (the DV) at various points in a single regime's development to show both

## Beyond Story-Telling

the conditions under which and pathways by which regime change leads to behavior change (Young 1999). Similarly, Parson examines different periods in the stratospheric ozone regime (a single case) to demonstrate that scientific assessments had decisive influence on international policy during the regime's formation but that technology assessments that drew heavily on private-industry expertise drove the rapid adaptation of the regime thereafter (2002, 2003). The quality of a QER study depends on whether the researcher selects her cases and observations in ways that allow her convincingly to evaluate its hypotheses.

### Selecting Cases and Observations

Careful case and observation selection that reflects prior theory and research lies at the heart of QER research in IEP. Indeed, we believe that the choices of whether to study one or more than one case, which case or cases to study, and how many observations of each case to make are as important as which questions are asked and which variables are included. In a QER study of IEP, the number and appropriateness of case selection and observation selection determine, even before research has begun, whether the researchers will be able to make a convincing argument about the hypotheses being tested. The researchers must choose which, and how many, cases and observations to make in light of how many IVs they seek to analyze. Adding another case or observation can either allow analysis of an additional IV or strengthen the analysis of the IVs already selected for study. We believe that IEP research progresses faster from studies that provide strong support for the hypotheses they evaluate rather than those that evaluate many hypotheses. Good scholarship is defined by the quality of analysis, not the quantity of analyses.

The quality of a QER study and its ability to identify specific effects of one or more independent variables is enhanced by the criteria used to select cases and observations. Cases and observations should be selected to ensure they provide data appropriate to the hypotheses being evaluated; the primary IV of interest varies across observations, while other IVs do not; and the number of observations exceeds the number of IVs that vary across those cases and observations. Assuming the quality of observations can be maintained, more observations are preferable to fewer. Observations also should be selected based on variation in the IVs of interest, "without regard

to the values of the dependent variable,” to avoid biasing the selection process in favor of the hypothesis but without selecting cases that exhibit no variation in the DV (King, Keohane, and Verba 1994, 142–46).

A particularly common threat to case selection in IEP arises from the understandable and appropriate desire to contribute to current policy debates. This desire produces an independent (but often inappropriate) precommitment to studying a particular case. Unfortunately, initiating research by selecting a case because of substantive interests or letting this case dictate the research question often produces results with neither theoretical nor policy value. Frequently, theoretically hot questions cannot be answered using politically hot cases. Thus, although scholars and practitioners are currently quite interested in regime effectiveness, studies of headline issues—climate change, desertification, deforestation—cannot provide useful insights into how to design regimes effectively because these cases have not yet produced data that could be used in service of this question. Likewise, QER studies that sought to evaluate the influence of free trade on environmental quality could only begin to use NAFTA (signed in 1993) as a case in the late 1990s, after NAFTA-related data became available (Thompson and Strohm 1996). Instead of analyzing hot issues, scholars concerned with contributing to current policy debates are better served by selecting historical cases that provide internally valid results but that, by intention, generalize particularly well to those cases of policy concern.

Isolating one IV's influence on a DV from another requires selecting observations, whether multiple observations of a single case or single observations of multiple cases, in which the IV of interest varies while the other IVs that may also influence the DV do not. Drawing convincing causal inferences requires, first of all, an unbiased assessment of the association between independent and dependent variables across observations—if there is no association between a hypothesized IV and a specific DV, change in this IV cannot be a cause of change in the DV. As a basic principle, researchers must include at least one observation for every value of the independent variable. Consider a hypothesized relationship between an IV that can have the values A and not-A, and a DV that can have the values B and not-B. To infer from the data that “A caused B,” the minimum requirements would be to show that (1) in one observation, B is present when A is present, (2) in a similar observation (i.e., *ceteris paribus*), B is absent when A is absent, and (3) A preceded B in time (Fearon 1991). This

has implications for whether one needs multiple cases or can conduct the analysis with multiple observations from a single case: evaluating the “culture-regime participation” proposition mentioned earlier most likely requires at least two cases of different cultures, since culture, however categorized, is not something that is likely to vary quickly over time.

Convincing inferences by the QER researcher are fostered by selecting observations that show other potentially explanatory variables did not vary (i.e., control variables). Analyzing a single environmental issue as composed of several observations, either by observing variables in different time periods or in different subissue areas, provides a strong argument that most variables did not vary across the comparison. This is particularly useful when an alternative explanatory variable, had it varied across the observations, would have explanatory priority. For example, Mitchell (1994) compared oil company compliance with two different but contemporaneous rules within a single international regime regulating oil pollution. This strategy ensured that economic influences (particularly oil prices, which would have trumped regime influences) had equivalent effects over compliance with both rules and could therefore be excluded as explanations of the observed differences in compliance with the different rules.

Or, consider the proposition that trade liberalization causes convergence of environmental standards across countries. In this proposition, the relevant class is all cases of “trade liberalization,” and an observation is defined by a pair of countries with some given level of trade. To evaluate the effect of trade liberalization requires comparing at least two such observations, one of which exhibits more liberalization than the other. The proposition predicts that the environmental standards of the countries with more liberalization among them will look more similar (their environmental standards will have converged more) than those with less liberalization. But, since two countries’ environmental policies may converge because they face the same transboundary environmental problem (rather than because of trade liberalization), the observations selected should ensure the environmental problem is local and varies significantly in magnitude across the two countries. The comparison then can occur between observations of a single pair of countries before and after an increase in liberalization, observations of two pairs of countries that have different levels of liberalization at a given point in time, or observations of both types. The first approach keeps most country-specific variables constant, the second keeps most time-specific variables constant, and the third allows an analysis that takes

## Models, Numbers, and Cases

account of the influence of both country-specific and time-specific variables. Evidence of greater policy similarity in addressing local environmental problems in liberalized settings than in nonliberalized ones supports the proposition. The first (change over time) approach allows evaluation of whether initially disparate national policies became more similar after liberalization, allowing us to concentrate our search for explanations among those, presumably few, variables that also changed at approximately the same time as the liberalization.

The notion that comparisons occur across observations, rather than cases, helps facilitate causal evaluations. Whether or not the real world of IEP provides observations that meet the criteria just mentioned, careful counterfactual analysis can prompt the collection of additional evidence that strengthens the conclusions of the study (Tetlock and Belkin 1996, 16–37; Fearon 1991; Dessler 1991). The analyst seeks to carefully identify whether and how the world would have been different had the independent variable of interest had a different value. For example, case studies on the effectiveness of international environmental regimes often ask questions such as how sulfur dioxide emissions in Western Europe would have evolved if the regime on long-range transboundary air pollution had not been established (Levy 1995).

A study's internal and external validity are also strengthened by seeking out "hard cases" in which the values of theoretically important control variables (CVs) are "distinctly unfavorable" to the hypothesis being tested (Young 1992, 165). Selection of observations should not only hold potentially explanatory variables constant, but should do so at values that make it unlikely the DV would have the value predicted by the IV of interest. This approach makes it less likely that evidence supporting the hypothesized relationship will be found but ensures such evidence, if found, provides stronger support of that hypothesis, since it was found true in an unlikely context (see, for example, Mitchell 1994; Haas 1990, 214; Young 1989). Also, such an approach strengthens the generalization of the case to other cases where the control variables had values more favorable to the IV of interest's influence on the DV.

Selecting cases and observations using these guidelines requires initial knowledge of possible cases. Gaining such knowledge usually requires discussions with other scholars, scanning primary and secondary literatures, and sometimes conducting one or more pilot studies. Following the guidelines discussed here can be painful: analysts often must drop cases to which

## Beyond Story-Telling

they are emotionally attached and in whose analysis considerable and unretrievable time and resources have already been invested, add cases or observations not yet investigated, or completely reanalyze cases and observations already studied. Yet, these steps are crucial to the eventual quality of the findings.

### Linking Data to Propositions

To collect data, the analyst must identify appropriate proxies of the study's theoretical constructs (Levy, Young, and Zürn 1995; Bernauer 1995; Young 1992; Dessler 1991, 339). Variables must be operationalized to correspond well to the relevant theoretical constructs (construct validity). General theories seek to explain outcomes across an array of cases and define variables nonspecifically. Yet, constructing falsifiable predictions requires defining variables in ways that help empirical identification of the values of and relationships among variables. For example, Sprinz and Vaahoranta conceive of interests in international environmental policy as consisting of a state's "environmental vulnerability" and "abatement costs," operationalizing them for ozone depletion as skin cancer rates and CFC consumption (1994, 89). Although these proxies fail to capture the conceptual richness of ecological vulnerability or abatement costs, they correspond reasonably well to these concepts and can be readily observed. Given the difficulties of operationalizing complex conceptual constructs and variables (Homer-Dixon 1996), it often helps to identify and collect data on a variety of indicators of a variable. Even if one has low confidence that any one of multiple indicators captures the true value of a conceptual variable, if all those indicators point in the same direction for that variable it can strengthen the confidence in the study's claims.

Examining different proxies may also lead to new theoretical insights. Victor, Raustiala, and Skolnikoff's (1998) analysis of various indicators of "implementation" documented that formal legal compliance with environmental treaties is often high whereas behavioral effectiveness, measured as whether a regime alters actor behavior, is low. By contrast, they found informal commitments associated with less compliance but more effectiveness, opening up an avenue for innovative explanations of variation across different indicators for effectiveness. QER studies also can identify variables initially considered unimportant and excluded from the original research design. A healthy tension can exist between the structure imposed

## Models, Numbers, and Cases

by delineating the relationships among IVs, DVs, and CVs before the study begins and, when linking data to propositions, an open-mindedness to factors not captured within the boundaries of this initial structure. For example, the finding that capacity conditions environmental treaty effectiveness was an inductive conclusion from the evidence rather than an affirmation of a deductive hypothesis in Haas, Keohane, and Levy's book (1993).

Clear definitions and operationalizations of variables allow the researcher to gather data representing the values for each IV, CV, and DV. Sources of data include the primary and secondary literature, documents, electronic databases, structured or open interviews and surveys, or direct or participant observation. QER research often combines qualitative and quantitative data. For example, in examining whether a particular international agreement reduced pollution levels, a researcher may seek quantitative and qualitative evidence of changes in pollution levels, changes in pollution control technologies, exogenous economic trends that might produce changes in pollution levels, as well as information on decision makers' and activists' views on whether and why pollution levels changed over time and how the agreement contributed to these reductions (Bernauer 1996).

### Examining Explanatory Pathways

Procedures for analyzing qualitative data are less well-established and less well-accepted than procedures for quantitative research. That said, the simple rules of systematically comparing predicted and observed values of the DV, qualitatively estimating associations between key variables, constructing empirical narratives to evaluate theoretical arguments more directly, and assessing rival hypotheses promotes solid inference (see, for example, Ragin 1994). The best approach for using empirical evidence to evaluate a theory is to identify clearly what the theory under investigation would predict about that DV's value given the values of the IVs and CVs for each observation, temporarily ignoring any knowledge about the value of the DV. The failure of observed DV values to match predicted DV values indicates that the theory fails to explain the variation across the observations in question, that it deals with a necessary but not sufficient condition for the DV to have the predicted value, or that it lacks explanatory power more generally.<sup>4</sup>

## Beyond Story-Telling

A table of the values of each variable for each observation imposes beneficial rigor on the analysis. Each row constitutes an observation with columns showing the values of each IV and CV, the value predicted for the DV given the values of the IVs and CVs, and the observed values of the DV. Such a table makes explicit the researcher's interpretation of what theory predicts in each observation, and it clarifies whether the IVs and DV covary as predicted. If they do, then the analyst needs to determine whether the covariation is evidence of a causal relationship or simply spurious covariation, which may involve including the values of other CVs in the chart to check for their covariation with the DV. Demonstrating that control variables actually were held constant (or had values that theory predicts would cause the DV to have a different value than that observed) refutes rival hypotheses that those variables caused the observed variation in the DV. If the control variables assumed as constant actually varied, the researcher should reevaluate her analytic strategy. Gathering more observations allows evaluation of whether the IV and/or (what was originally considered a) control variable explain the DV's variation. In this and other instances, change in CVs requires careful attention to ensure that a presumed control variable is not an important cause of change in the DV.

Consider a study evaluating how scientific consensus influences the outcome of international environmental negotiations. The independent variable might be defined as "level of scientific consensus on the causes of an environmental problem" with potential values of high or low; and the dependent variable as "negotiation of international environmental agreement" with potential values of yes or no. Evaluating this hypothesis requires—as a starting point—comparing one observation exhibiting high consensus regarding an environmental problem's causes with another exhibiting low consensus. Four types of results might arise: (1) a negotiated agreement in the first but not the second; (2) a negotiated agreement in the second but not the first; (3) negotiated agreements in both; and (4) no negotiated agreement in either. In results (3) and (4), no association between the two variables appears to exist. Result (1) provides initial support that scientific consensus may facilitate negotiation of an agreement, while result (2) provides initial support that scientific consensus may inhibit negotiation of an agreement.

Having found covariation between the IV of interest and the DV (with CVs held constant), the researcher should identify how the IV explains variation in the DV. Such a narrative will usually combine theoretical rea-

soning with empirical evidence on hypothesized “causal mechanisms” that connect IVs and the DV. As Alexander George notes, this involves subjecting the observations “in which that correlation appears to more intensive scrutiny, as the historian would do, in order to establish whether there exists an *intervening process*, that is, a causal nexus, between the independent variable and the dependent variable” (1979, 46). QER research has an advantage over quantitative methods in this regard, since it allows disaggregated and in-depth analysis of such mechanisms (Dessler 1991, 352). Detailed process tracing is more than mere storytelling (George 1979). Examining causal pathways helps demonstrate a linkage between an IV and a DV by taking advantage of the fact that a theory regarding an IV usually has several, and often many, observable implications beyond simply the value of the DV. Process tracing involves evaluating available evidence to see if it supports these additional implications that the IV explains change in the DV. Thus, to evaluate whether a given case supports Homer-

Dixon’s claim that environmental degradation causes acute conflict requires demonstrating not merely that environmental degradation was followed by acute conflict but that environmental degradation was followed by natural resource scarcity, which was followed by increased migration, which was followed by exacerbated rivalries between immigrants and indigenous populations, which was followed by the observed conflict (1991, 85–86; 1996).

Quantitative methods exist to evaluate such pathways, but collecting this level of information for more than a few cases often exceeds available resources. QER methods therefore complement quantitative methods. Internal validity is enhanced if alternative explanations are considered and found “less consistent with the data and/or less supportable by available generalizations” (George 1979, 57–58). Standard theoretical explanations of variation in the DV provide a list of likely suspects for these rival hypotheses. As already noted, careful selection of cases and observations should have eliminated many of these variables by holding them constant. The researcher should evaluate explicitly whether remaining rival variables provide a better or simpler explanation of the observed variation in the DV. She also should give the benefit of the doubt to these rival explanations if those explanations have strong previous support. For example, one explanation of why governments negotiated and signed the Montreal Protocol in 1987 contends that scientific consensus, stimulated mainly by the dis-

covery of the ozone hole, was the proximate cause of negotiating success (Haas 1992). Making this claim requires refuting the rival claim that governments completed the agreement in response to Dupont (the world's leading producer of CFCs) and later other producers removing their objections to the agreement for economic, not scientific or environmental, reasons (Parson 2003). Notably, the rival arguments can be compared based on their predictions of when, as well as whether, agreement would be reached. Similarly, Bernauer and Moser (1996) evaluated whether the Rhine protection regime caused decreased pollution levels by developing several plausible narratives leading from the presumed cause to the observed outcome. Process tracing via open interviews with decision makers, analysis of pollution and regulatory data, and review of the secondary literature showed that transboundary regulation probably caused only a fraction of the pollution reduction. Interviews in particular showed that a secular decline in heavy industries and coal mining along the Rhine, and a general greening of domestic politics, also appeared to correlate with pollution reductions, making them more likely and plausible explanatory factors. Their careful evaluation of rival hypotheses allowed these analysts to avoid inaccurately attributing causal power to the international regime.

If the observations and cases selected allow too many plausibly explanatory variables to vary—a situation QER researchers often find themselves in—they face the difficult task of partitioning explanatory power between two or more IVs. Evaluating the implications of competing explanatory variables against available evidence often helps here. For example, both epistemic community and industry pressure arguments seem to explain the content of the Montreal Protocol. Closer consideration, however, suggests that the inclusion of various ozone depleting substances in the Montreal Protocol reflected the epistemic community's influence, while the Protocol's focus on alternatives involving new chemicals rather than nonchemical technologies reflected industry's influence. With sufficient observations, the analyst can examine whether the covariance of one IV with the DV is more consistent across different subsets of observations than the covariance of the other IV with the DV. Although the number of observations in a study is usually too small to build a strong argument to choose between the two alternatives, such an analysis makes the most of the available data.

The foregoing discussion highlights a unique risk of QER research on IEP: no fallback position exists if no evidence of the hypothesized causal

## Models, Numbers, and Cases

relationship is found. The need to exclude observations that vary in explanatory variables other than the IV of interest runs the risk that, if the IV of interest does not explain variation in the DV, the researcher will have to (1) examine new cases to see if, under other conditions, the hypothesized relationship does exist or (2) collect new data from the cases studied to determine whether other variables previously considered unimportant actually contributed to the observed variation in the DV.<sup>5</sup> Many scholars can recount studies involving carefully selected cases and observations that simply failed to allow development of a compelling account that could exclude other factors as explanations of variation in the DV.

## Generalizing to Other Cases

Finally, QER researchers on IEP should evaluate whether their findings are relevant to cases not investigated.<sup>6</sup> Assessing external validity forges links with broader debates. If the cases and observations studied differ in significant ways from most cases in a class, accurate generalization becomes difficult. Even if they can be generalized, convincing policymakers to adopt or reject a particular policy (or scholars to accept or reject a particular theory) requires demonstrating that the conditions in the cases studied are sufficiently similar to those in the targeted policy area to warrant the expectation that the same explanatory relationships will operate there.

It can be argued that case studies provide less foundation for generalization than quantitative analyses. In a quantitative, large-*N* study, the pursuit of internal validity usually also ensures that the results apply to a range of cases, both because they apply to the range of cases included in the study and because when findings are derived from observations and cases that exhibit considerable variation in other variables, it is more likely that they apply to a broader set of cases. High-quality quantitative analysis can identify both the average association of the IV with the DV over a range of CV values as well as interactive effects between the IV and various CVs (Mitchell 2002a). In contrast, most QER studies document the association between an IV and a DV for only a single value of any given CV, a CV that is likely to be different in many cases to which the researcher might want to generalize.

Consider a quantitative and a qualitative study of how NGO activism influences the chances of regime formation. Imagine that the only variable

that theory suggests needs to be controlled for is the existence of a hegemon (Young 1983). Assuming the availability of sufficient cases with and without NGO activism, with and without hegemon, and with and without regimes, the quantitative analyst could identify the average impact NGOs had on regime formation and might be able to determine whether this impact was larger when a hegemon was absent and smaller when a hegemon was present. In contrast, the qualitative analyst, knowing that hegemony would likely confound causal inference, would seek to hold hegemony constant, selecting “hard cases” where hegemon was absent since they are alleged to facilitate regime formation. The qualitative study should find the same strong influence of NGOs on regime formation as the quantitative study found among cases in which hegemon was absent. The qualitative study would have developed a clearer sense of exactly how the NGOs fostered regime formation. Although the analyst should not bluntly claim that “NGOs play a major role in regime formation,” she could claim that “these cases support the view that NGOs facilitate regime formation when no hegemon is available.” Having made this argument with these observations, she might then go on to explore whether NGOs are likely to facilitate regime formation when a hegemon is available or if they have no real influence in that situation. Had she selected cases where hegemon (which promote regime formation) were present, however, both the internal validity and external validity of the study would be more limited.

Given internally valid findings, analysts can increase external validity in three ways. First, they can add cases and observations in which CVs have different values. Second, they can select hard cases (see previous discussion as well as Bennett, chap. 2, and Odell, chap. 3, this vol.). Third, they can select cases and observations that take advantage of the types of variation examined in past studies. By literal or theoretical replication of other cases, choosing cases to evaluate how different values of control variables influence the relationship of the IV and the DV, the researcher can contribute to a progressive research program that collectively establishes generalizable results through joint analysis of a larger set of cases (Yin 1994). In sum, the researcher should be cautious in generalizing. Assessing external validity requires determining how the values at which CVs were held constant affected the influence of the IV on the DV and on how well the values of those CVs match other cases and observations in the class of cases being studied.

## Conclusion

More of our knowledge of IEP to date has come from QER studies than from quantitative studies or formal studies, and that is likely to remain true for the foreseeable future. Yet, methodologies for drawing inferences from small numbers of cases and observations remain underdeveloped and too rarely appear on political science syllabi. The plethora of poorly constructed QER studies in the IEP literature testifies to the difficulty of designing and conducting them well. To address this, we have outlined six steps crucial to rigorous QER research in IEP. We recognize that other qualitative approaches have much to offer the study of IEP and urge others to provide guidelines for good research within those traditions.

Good IEP research in what we have called the qualitative, explanatory, rationalist (or QER) tradition begins by identifying an important and compelling research question. The analyst then transforms this question into testable hypotheses, explicitly delineating variables and their possible values. Having done this, choices about the number and type of cases and observations provide a crucial, but all too often ignored or poorly implemented, basis for identifying associations between independent and dependent variables, drawing causal inferences from such findings, and generalizing those findings to other cases. Cases and observations should be selected to facilitate the researcher's urge to develop or evaluate a theory, avoiding the temptation to explore politically hot or personally interesting cases unless they also can answer theoretically important questions. QER research requires sufficient familiarity with existing theory to identify the independent and control variables considered most important to variation in the dependent variable. Simple tables can help identify the degree to which observed variation in the DV corresponds (*ceteris paribus*) to variation predicted by theory. Process tracing of other observable linkages between the IV and DV can strengthen claims that those relationships are causal and not just correlational. Finally, the researcher should clearly identify the class of cases to which the findings can (and cannot) be appropriately generalized.

In QER research on IEP, systematic evaluation of hypothesized relationships must be restricted to very few IVs and CVs. Even with a solid research design, comparing the effects of more than one IV on the DV is much harder than with multivariate statistical analysis. Moreover, explanations derived from QER studies are often difficult to generalize. The fre-

## Beyond Story-Telling

quently encountered view that QER studies are an “easy way out” for students and scholars deterred by the technicalities of statistics is highly erroneous. As the difficulties discussed here clarify, deriving valid and interesting inferences from a QER study requires at least as much care and thought as doing so from a quantitative analysis. Those undertaking QER studies are well advised to follow Mark Twain’s admonition to “put all your eggs in the one basket and—WATCH THAT BASKET” (1894, 15). The procedures outlined here can reduce the risks. We believe that researchers attentive to these steps can improve the rigor of their work, contributing to the growing effort to understand how international environmental problems arise; when, why, and how they can be resolved; and how we can better manage our social, political, and economic behaviors to have fewer detrimental impacts on the environment.

## Recommended Readings

- Brown Weiss, E., and H. K. Jacobson, eds. 1998. *Engaging Countries: Strengthening Compliance with International Environmental Accords*. Cambridge: MIT Press.
- Haas, P. M., R. O. Keohane, and M. A. Levy, eds. 1993. *Institutions for the Earth: Sources of Effective International Environmental Protection*. Cambridge: MIT Press.
- Miles, E. L., A. Underdal, S. Andresan, J. Wettestad, J. B. Skjaerseth, and E. M. Carlin, eds. 2001. *Environmental Regime Effectiveness: Confronting Theory with Evidence*. Cambridge: MIT Press.
- Victor, D. G., K. Raustiala, and E. B. Skolnikoff, eds. 1998. *The Implementation and Effectiveness of International Environmental Commitments*. Cambridge: MIT Press.
- Young, O. R., ed. 1999. *Effectiveness of International Environmental Regimes: Causal Connections and Behavioral Mechanisms*. Cambridge: MIT Press.

## Notes

This chapter is a substantially revised version of Ronald B. Mitchell and Thomas Bernauer, “Empirical Research on International Environmental Policy: Designing Qualitative Case Studies,” *Journal of Environment and Development* 7 (1): 4–31, March 1998. The authors would like to express their appreciation to Detlef Sprinz and Yael Wolinsky-Nahmias, as well as to Peter Haas, Andrew Bennett, Harold Jacobson, Peter Moser, Roy Suter, Arild Underdal, and two anonymous reviewers of the *Journal of Environment and Development* for extremely helpful comments on earlier drafts of this article.

## Models, Numbers, and Cases

Ronald Mitchell's work was supported by a Sabbatical Fellowship in the Humanities and Social Sciences from the American Philosophical Society.

1. These steps draw extensively from the work of Tetlock and Belkin (1996); King, Keohane, and Verba (1994); George (1979); Eckstein (1975); and other scholars of qualitative research methodology.

2. Members of dissertation committees as well as discussants on panels at scientific conferences regularly ask questions such as "And what is this a case of?" Too frequently, however, the answers reveal a failure to consciously address this question early on in the research process.

3. George and Bennett (1997) have argued that case studies can be particularly useful for developing "typological theory" in which the case study helps to establish the different categories of phenomena as well as their causal relationships.

4. Of course, the first two of these possibilities can and should have been addressed by careful selection of cases to allow more determinate testing of the theory chosen for study.

5. The QER researcher must designate an independent variable of interest at the outset of the study, then face the problem of having "no findings" if variation in that independent variable is not associated with variation in the dependent variable. The quantitative researcher, by contrast, can select cases and observations that exhibit considerable variation in all the independent variables, developing an explanatory argument around whichever independent variables show strong associations with the dependent variable.

6. Many constructivists argue that the importance of contextual contingency demands a healthy skepticism about even limited generalizations.

## References

- Barnett, M. N., and M. Finnemore. 1999. The Politics, Power, and Pathologies of International Organizations. *International Organization* 53 (4): 699–732.
- Bernauer, T. 1995. The Effect of International Environmental Institutions: How We Might Learn More. *International Organization* 49 (2): 315–77.
- . 1996. Protecting the Rhine River against Chloride Pollution. In *Institutions for Environmental Aid: Pitfalls and Promise*, edited by R. O. Keohane and M. A. Levy. Cambridge: MIT Press.
- Bernauer, T., and P. Moser. 1996. Reducing Pollutions of the Rhine River: The Influence of International Cooperation. *Journal of Environment and Development* 5 (4): 391–417.
- Biersteker, T. 1993. Constructing Historical Counterfactuals to Assess the Consequences of International Regimes: The Global Debt Regime and the Course of the Debt Crisis of the 1980s. In *Regime Theory and International Relations*, edited

## Beyond Story-Telling

- by V. Rittberger. New York: Oxford University Press.
- Breitmeier, H., M. A. Levy, O. R. Young, and M. Zürn. 1996. *The International Regimes Database as a Tool for the Study of International Cooperation*. Laxenburg, Austria: International Institute for Applied Systems Analysis.
- Brown Weiss, E., and H. K. Jacobson. 1998. *Engaging Countries: Strengthening Compliance with International Environmental Accords*. Cambridge: MIT Press.
- Clapp, J. 1998. The Privatization of Global Environmental Governance: ISO 14000 and the Developing World. *Global Governance* 4 (3): 295–316.
- Dawson, J. 1996. *Eco-Nationalism*. Durham: Duke University Press.
- DeSombre, E. R. 2000. *Domestic Sources of International Environmental Policy: Industry, Environmentalists, and U.S. Power*. Cambridge: MIT Press.
- Dessler, D. 1991. Beyond Correlations: Toward a Causal Theory of War. *International Studies Quarterly* 35 (3): 337–55.
- Eckstein, H. 1975. Case Study and Theory in Political Science. In *Handbook of Political Science*, vol. 7, *Strategies of Inquiry*, edited by F. Greenstein and N. Polsby. Reading, MA: Addison-Wesley.
- Fearon, J. D. 1991. Counterfactuals and Hypothesis Testing in Political Science. *World Politics* 43 (2): 169–95.
- George, A. L. 1979. Case Studies and Theory Development: The Method of Structured, Focused Comparison. In *Diplomacy: New Approaches in History, Theory, and Policy*, edited by P. G. Lauren. New York: Free Press.
- George, A. L., and A. Bennett. 1997. Developing and Using Typological Theories in Case Study Research. International Studies Association Conference, Toronto.
- . Forthcoming. *Case Studies and Theory Development*. Cambridge: MIT Press.
- Haas, P. M. 1990. *Saving the Mediterranean: The Politics of International Environmental Cooperation*. New York: Columbia University Press.
- . 1992. Banning Chlorofluorocarbons. *International Organization* 46 (1): 187–224.
- Haas, P. M., R. O. Keohane, and M. A. Levy, eds. 1993. *Institutions for the Earth: Sources of Effective International Environmental Protection*. Cambridge: MIT Press.
- Helm, C., and D. Sprinz. 2000. Measuring the Effectiveness of International Environmental Regimes. *Journal of Conflict Resolution* 44 (5): 630–52.
- Homer-Dixon, T. 1991. On the Threshold: Environmental Changes as Causes of Acute Conflict. *International Security* 16 (2): 76–116.
- . 1996. Strategies for Studying Causation in Complex Ecological Political Systems. *Journal of Environment and Development* 5 (2): 132–48.
- Keohane, R. O., and M. A. Levy, eds. 1996. *Institutions for Environmental Aid: Pitfalls and Promise*. Cambridge: MIT Press.
- King, G., R. O. Keohane, and S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Lakatos, I. 1970. Falsification and the Methodology of Research Programmes. In

## Models, Numbers, and Cases

- Criticism and the Growth of Knowledge*, edited by I. Lakatos and A. Musgrave. Cambridge: Cambridge University Press.
- Levy, M. A. 1995. International Cooperation to Combat Acid Rain. In *Green Globe Yearbook: An Independent Publication on Environment and Development*, edited by H. O. Bergesen, G. Parmann, and Ø. B. Thommessen. Oxford: Oxford University Press.
- Levy, M. A., O. R. Young, and M. Zürn. 1995. The Study of International Regimes. *European Journal of International Relations* 1 (3): 267–330.
- Lipschutz, R. D. 1991. Bargaining among Nations: Culture, History, and Perceptions in Regime Formation. *Evaluation Review* 15 (1): 46–74.
- Lipschutz, R. D., and K. Conca, eds. 1993. *The State and Social Power in Global Environmental Politics*. New York: Columbia University Press.
- Meyer, J. W., D. J. Frank, A. Hironaka, E. Schofer, and N. B. Tuma. 1997. The Structuring of a World Environmental Regime, 1870–1990. *International Organization* 51 (4): 623–29.
- Miles, E. L., A. Underdal, S. Andresen, J. Wettestad, J. B. Skjærseth, and E. M. Carlin, eds. 2000. *Explaining Regime Effectiveness: Confronting Theory with Evidence*. Cambridge, MA: MIT Press.
- Miles, E. L., A. Underdal, S. Andresen, J. Wettestad, J. B. Skjaerseth, and E. M. Carlin, eds. 2001. *Environmental Regime Effectiveness: Confronting Theory with Evidence*. Cambridge: MIT Press.
- Mitchell, R. B. 1994. *Intentional Oil Pollution at Sea: Environmental Policy and Treaty Compliance*. Cambridge: MIT Press.
- . 2002a. A Quantitative Approach to Evaluating International Environmental Regimes. *Global Environmental Politics* 2 (4): 58–83.
- . 2002b. International Environment. In *Handbook of International Relations*, edited by W. Carlsnaes, T. Risse, and B. Simmons. Thousand Oaks, CA: Sage.
- . In press. International Environmental Agreements: A Survey of their Features, Formation, and Effects. *Annual Review of Environment and Resources* 28.
- Mitchell, R. B., W. C. Clark, D. W. Cash, and F. Alcock, eds. Forthcoming. *Global Environmental Assessments: Information, Institutions, and Influence*. Cambridge: MIT Press.
- O'Neill, K. 2000. *Waste Trading among Rich Nations*. Cambridge: MIT Press.
- Parson, E. A. 2002. The Technology Assessment Approach to Climate Change. *Issues in Science and Technology* 18 (4): 65–72.
- . 2003. *Protecting the Ozone Layer: Science and Strategy*. Oxford: Oxford University Press.
- Popper, K. 1968. *The Logic of Scientific Discovery*. New York: Harper and Row.
- Princen, T. 1996. The Zero Option and Ecological Rationality in International Environmental Politics. *International Environmental Affairs* 8 (2): 147–76.

## Beyond Story-Telling

- Princen, T., and M. Finger. 1994. *Environmental NGOs in World Politics: Linking the Local and the Global*. New York: Routledge.
- Ragin, C. C. 1994. *Constructing Social Research: The Unity and Diversity of Method*. Thousand Oaks, CA: Pine Forge Press.
- Ruggie, J. G. 1998. What Makes the World Hang Together? Neo-Utilitarianism and the Social Constructivist Challenge. *International Organization* 52 (4): 855–85.
- Sand, P. H., ed. 1992. *The Effectiveness of International Environmental Agreements: A Survey of Existing Legal Instruments*. Cambridge, UK: Grotius.
- Schreurs, M. A., and E. Economy, eds. 1997. *The Internationalization of Environmental Protection*. Oxford: Oxford University Press.
- Sjostedt, G., ed. 1993. *International Environmental Negotiation*. Newbury Park, CA: Sage.
- Social Learning Group, ed. 2001a. *Learning to Manage Global Environmental Risks*. Vol. 1, *A Functional Analysis of Social Responses to Climate Change, Ozone Depletion, and Acid Rain*. Cambridge: MIT Press.
- . 2001b. *Learning to Manage Global Environmental Risks*. Vol. 2, *A Comparative History of Social Responses to Climate Change, Ozone Depletion, and Acid Rain*. Cambridge: MIT Press.
- Sprinz, D., and T. Vaahoranta. 1994. The Interest-Based Explanation of International Environmental Policy. *International Organization* 48 (1): 77–105.
- Stinchcombe, A. L. 1968. *Constructing Social Theories*. New York: Harcourt, Brace, and World.
- Susskind, L. E., E. Siskind, and J. W. Breslin, eds. 1990. *Nine Case Studies in International Environmental Negotiation*. Cambridge: MIT-Harvard Public Disputes Program.
- Tetlock, P. E., and A. Belkin, eds. 1996. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton: Princeton University Press.
- Thompson, P., and L. A. Strohm. 1996. Trade and Environmental Quality: A Review of the Evidence. *Journal of Environment and Development* 5 (4): 363–88.
- Twain, M. 1894. *The Tragedy of Pudd'nhead Wilson; and, the Comedy, Those Extraordinary Twins*. Hartford: American Publishing Company.
- Underdal, A. 2001. Conclusions: Patterns of Regime Effectiveness. In *Environmental Regime Effectiveness: Confronting Theory with Evidence*, edited by E. L. Miles, A. Underdal, S. Andresen, J. Wettestad, J. B. Skjaereth, and E. M. Carlin. Cambridge: MIT Press.
- Victor, D., K. Raustiala, and E. B. Skolnikoff. 1998. *The Implementation and Effectiveness of International Environmental Commitments*. Cambridge: MIT Press.
- Wapner, P. 1996. *Environmental Activism and World Civic Politics*. Albany: State University of New York Press.

## Models, Numbers, and Cases

- Wettestad, J. 1999. *Designing Effective Environmental Regimes: The Key Conditions*. Cheltenham, UK: Edward Elgar.
- Yin, R. K. 1994. *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage.
- Young, O. R. 1983. Regime Dynamics: The Rise and Fall of International Regimes. In *International Regimes*, edited by S. D. Krasner. Ithaca: Cornell University Press.
- . 1989. *International Cooperation: Building Regimes for Natural Resources and the Environment*. Ithaca: Cornell University Press.
- . 1992. The Effectiveness of International Institutions: Hard Cases and Critical Variables. In *Governance without Government: Change and Order in World Politics*, edited by J. N. Rosenau and E.-O. Czempiel. New York: Cambridge University Press.
- . 1998. *Creating Regimes: Arctic Accords and International Governance*. Ithaca: Cornell University Press.
- , ed. 1999. *Effectiveness of International Environmental Regimes: Causal Connections and Behavioral Mechanisms*. Cambridge: MIT Press.
- . 2001. Inferences and Indices: Evaluating the Effectiveness of International Environmental Regimes. *Global Environmental Politics* 1 (1): 99–121.
- Young, O. R., and G. Osherenko, eds. 1993. *Polar Politics: Creating International Environmental Regimes*. Ithaca: Cornell University Press.