

Empirical Research on International Environmental Policy: Designing Qualitative Case Studies

RONALD MITCHELL
THOMAS BERNAUER

Although empirical research on international environmental politics and policy (IEP) will, by necessity, rely on qualitative case studies, this methodology remains less developed than quantitative procedures. Analysts using qualitative case studies to evaluate and generalize causal inferences can improve their research by following six key research steps. Selecting cases carefully, drawing appropriate causal inferences, and addressing the tension between specificity and generalizability prove particularly important. Analysts should develop theoretically meaningful propositions before selecting cases. Theoretically and practically interesting questions often cannot be answered with politically "hot" cases. Drawing internally valid causal inferences requires clearly defining and measuring dependent, independent, and control variables and selecting cases to control for exogenous variables. Focusing on "hard cases" and explicitly analyzing rival hypotheses produce stronger causal inferences. Analysts should give precedence to internal validity over external validity in their findings. Examples drawn from several environmental issue areas illustrate the method and criteria.

The demands of a case-study on a person's intellect, ego, and emotions are far greater than those of any other research strategy.

Yin, 1994, p. 55

Social scientists interested in understanding international environmental politics and policy (IEP) can employ a range of research strategies—from game theory and regression analysis to simulations and experiments to surveys, historical analysis, and case studies. Those engaged in efforts to increase our knowledge of the causal relationships underlying IEP can use qualitative methods, quantitative methods, or both. The two methods can be distinguished by the

AUTHORS' NOTE: An earlier version of this article was presented at the March 1997 annual meeting of the International Studies Association, Toronto, Canada.

Journal of Environment & Development, Vol. 7, No. 1, March 1998 4-31
© 1998 Sage Publications, Inc.

type of information used, the procedures for processing information, and the number of cases typically analyzed.

Quantitative methodologists seek to understand causal relationships in IEP by conceptualizing, measuring, and analyzing information about the real world by means of numerical data representing explicitly defined variables. They analyze these data via statistical procedures to compare a large number of cross-sectional or longitudinal observations (or cases, see below) with the aim of identifying potentially strong, nonrandom, correlations between explanatory (or independent) variables and effects (or dependent variables). Statistical algorithms use the noncorrelated variation among the independent variables (IVs) to isolate the extent to which variation in the dependent variable (DV) would correlate with variation in the IV if other variables had been held constant.

Qualitative methodologists interested in evaluating and generalizing causal inferences pursue the same goal. They also rely on explicitly defined variables but capture the values of these variables in words and analyze these data through other techniques. Qualitative methods are particularly common when few cases exist for the researcher to use in evaluating theoretical claims. Instead of isolating causal relationships through large numbers of cases and statistical procedures, qualitative methodologists evaluate such relationships by holding other variables constant through careful case selection.

This article focuses on that subset of all qualitative case studies in which the study is defined as an observational study using qualitative information, few cases, and nonstatistical procedures to derive causal inferences and generalize them to other cases. Although qualitative analysts are frequently accused of being "soft," imprecise, subjective, or even unscientific, quantitative analysts are often accused of ignoring important social phenomena because they cannot be quantified, of failing to capture complex variables with the simplified data they use, and of producing precise but unreliable or irrelevant results with sophisticated statistical techniques but data of poor quality. These controversies obscure the fact that both quantitative and qualitative methods of identifying causal relationships demand similar care in designing and conducting empirical research. In this article, we seek to translate and expand general principles of case study methodology into specific guidance and criteria for those investigating the causes of variation in international

environmental politics, simultaneously introducing major features of this methodology to those unfamiliar with it, while elaborating theoretical and practical points that we believe have received inadequate attention to date.

Even scholars convinced that case studies can contribute to progress in the field differ in their views of how they should be conducted and what they can accomplish. Scholars have undertaken case studies that are descriptive, prescriptive, predictive, constructivist, reflectivist, interpretivist, and positivist in form (Moon, 1975). These different approaches have produced important, and often heated, debates about whether it is desirable or even possible to derive causal inferences from case studies or to generalize from such studies to other situations. These debates are unlikely to be definitively resolved in the near future, nor do we aspire to do so here. We believe that the current variety of methodological approaches all contribute to improving our chances of understanding IEP. Our limited goal here, therefore, is to provide guidance to that subset of students and scholars who, at least some of the time, undertake essentially positivist, "variation-finding" inquiries attempting to draw causal inferences about the sources of the variation one observes in IEP in ways that allow them to generalize those inferences to other cases. We outline six practical steps for addressing the particular obstacles to drawing systematic, rigorous, and theoretically and empirically informative findings from causal qualitative research in this particular field of study. We are motivated by our belief that studies seeking to identify and generalize causal inferences in the realm of international environmental politics too frequently make methodological errors that lead to unnecessarily weak, or inaccurate, conclusions. Rather than highlight the frequent negative examples of such errors, we illustrate the article with positive examples of researchers who have avoided such errors or with hypothetical examples of how to avoid them. We seek to adapt, refine, and consolidate general principles of case study methodology so that they can be applied more effectively in the realm of IEP, and we pay particular attention to the importance of and methods for useful case selection.

Unlike quantitative techniques, causal qualitative analysis of a small number of cases facilitates investigation of the following: (a) important but difficult-to-quantify variables (such as power, interests, or leadership); (b) theoretically important, empirically rare,

or previously ignored cases; (c) innovative (but, by their nature, rare) international environmental policy strategies; and (d) causal, rather than merely correlational, relationships (Baldwin, 1979; Yin, 1994, pp. 38-44; Young, 1994, p. 117). In particular, it allows one to look for evidence that a specific IV influenced the DV via a hypothesized causal pathway. A more nuanced understanding of causal pathways strengthens the argument that a causal relationship exists.

In the next part of the article, we identify five criteria for assessing the quality of case studies. We go on to argue that these standards can be met only if one pays attention to six steps in the research process. We discuss potential difficulties researchers face in each of these steps and propose ways to overcome them. These problems and possibilities are illustrated with a variety of examples drawn from different areas of IEP.

Performance Criteria and Research Steps

The large and well-established literature on quantitative empirical research methods in the social sciences has only recently begun to be matched by a complementary literature on designing and carrying out systematic and rigorous qualitative research (George & Bennett, 1997; King, Keohane, & Verba, 1994; Ragin, 1994; Tetlock & Belkin, 1996; Yin, 1994, pp. 54-101). This literature 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 summarized in Table 2. We lay out these tasks in logical order even while recognizing that research rarely does, and often should not, proceed in such a linear fashion; high-quality research often requires an iterative path through these steps. But those conducting qualitative case study research should address explicitly each of these tasks to produce findings that compel, convince, and contribute.¹ The most critical features of the six steps pertain to the selection of cases, the drawing of causal inferences, and the tension between specificity and generalizability.

1. Obviously, these steps draw from the work of Tetlock and Belkin (1996), King, Keohane, and Verba (1994), George (1979), Eckstein (1975), and others who have developed case study methodology.

Table 1
Criteria for Case Study Research

<i>Criterion</i>	<i>Question</i>
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 population 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—for example, 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?

IDENTIFYING AN IMPORTANT THEORETICAL QUESTION

High-quality research—whether qualitative or quantitative—must start with an important research question. Innovative causal case study research should thus begin with a question that addresses existing theoretical debates in the relevant field, aims at causal relationships, and attends to current policy concerns (King et al., 1994, p. 15). To contribute to a larger theoretical enterprise, research efforts should target unresolved debates, untested theoretical claims including those about the merits of unassessed policy “innovations,” or previously uninvestigated relationships. Appropriate targeting of research questions requires a sophisticated understanding of existing theory and knowledge of existing empirical patterns.

Becoming familiar with existing theoretical literature allows the researcher to simultaneously identify important questions and frame the research to target a particular scholarly community. Avoiding descriptive literature reviews, the researcher should delineate how major schools of international relations theory—for example,

Table 2
Research Tasks

<i>Task</i>	<i>Description</i>
1. Identifying an important theoretical question	Innovative causal analyses of IEP frame questions or empirical puzzles so that they address existing theoretical debates in the field, are targeted at causal relationships, and relate to current policy concerns. A particularly productive way to frame research involves evaluating the explanatory power of competing hypotheses or theories against evidence from relatively few cases.
2. Developing hypotheses and identifying variables	Translate general research questions or puzzles into explicit hypotheses with independent, dependent, and control variables. The analyst should identify the possible values of each variable and the evidence that could falsify her hypotheses. Studies making a single causal claim and concentrating on particular explanatory variables often produce more valid and reliable results than those making numerous claims.
3. Selecting cases	The most productive causal research focuses on theory first (Tasks 1 and 2) and selects empirical cases later. As with quantitative studies, valid inferences in qualitative research require more cases than explanatory variables. A case is a phenomenon for which one observes a single value for each variable in a hypothesis. Careful case selection seeks to approximate experimental conditions by holding specific exogenous (control) variables constant. The analyst can strengthen her causal claims by selecting hard cases in which the values of control variables make it unlikely that the explanatory variable will produce the theoretically predicted value of the dependent variable.
4. Linking data to propositions	The analyst should operationalize variables to facilitate valid and reliable measurement. The methods of the research process should be transparent to, and reliably replicable by, other researchers. Measurement will ideally be based on several different but complementary operationalizations of the variables.

(continued)

Table 2 Continued

5. Examining correlations and causal pathways	The analyst should systematically assess whether explanatory and dependent variables correlate as predicted by theory and investigate whether identified correlations reflect causal relationships. Causal analysis must be based on evaluating the predicted and observed values of the dependent variable, the corresponding causal narratives, and potential rival hypotheses.
6. Generalizing to other cases	Closing the research cycle requires relating findings back to the broader theoretical questions that motivated the research. The researcher should critically assess how far the findings generalize to other cases. Because internal validity is a necessary precondition for external validity and because the researcher often must trade off these two goals, internal validity should take precedence over external validity. Careful case selection facilitates generalizing findings accurately across a larger set of conditions.

realism, institutionalism, critical theory, constructivism—have answered the central research issue. If the question has not been explicitly addressed before, the analyst should deduce likely answers from core principles of broader theoretical arguments. For example, although few scholars had addressed regime compliance until recently, predictions about compliance with international environmental regimes could be readily derived from the theoretical assumptions of major schools of international relations theory (Haas, 1990, chap. 2; Jacobson & Brown Weiss, 1995; Mitchell, 1994, chap. 2; Young, 1989, chap. 3).

The newness of scholarly interest in IEP may tempt the researcher to view a particular issue as undertheorized because no one has explicitly sought to explain the phenomena of interest. The analyst should resist this temptation, however, and start from the assumption that general international relations theory applies equally well to IEP as to other policy domains. If this assumption clearly proves ill founded—for example, if power and interests explain less variation in the outcomes one observes in the realm of IEP than in the security realm—this suggests that prevailing theories of interna-

tional relations are less generalizable than claimed, rather than that no relevant theory exists. Indeed, far too often, analysts fail to draw on existing theories that could shed valuable insight on the international environmental problem under study.

Empirical “puzzles” often provide a particularly useful way to frame research questions. By highlighting a contradiction between the outcomes predicted by theory and the environmental outcomes actually observed, they both engage the reader and provide grounding for theory. Such puzzles can derive from competing theoretical explanations for the same outcome; from theoretical predictions, the accuracy of which has not yet been evaluated; or from the absence of any theories purporting to explain a particular observed outcome. For example, the extensive research on nonstate actors and epistemic communities has been motivated, in part, by the observation that these nongovernmental actors appear to have more influence on international environmental outcomes than predicted by orthodox, state-centric, models of international relations (Haas, 1992b; Princen, Finger, & Manno, 1995; Wapner, 1995). Likewise, much of the research on the Montreal Protocol negotiations has been motivated by the apparent contradiction between the speed and extent of actual international cooperation and the strong predictions of realist theories of international relations and collective action theories that international cooperation in these circumstances should be particularly difficult (Breitmeier, 1996; Haas, 1992a).

Analyzing a puzzle that engages a stark debate about the causes of a particular outcome ensures that the findings will refute one theory and lend support to another (or will refute both), regardless of the particular empirical findings. The research community understandably values findings disconfirming existing theory more than those supporting existing theory. Thus, testing a theory that has no competing theory is riskier because the researcher cannot know beforehand whether the cases selected will disconfirm the theory; she may end up with results that fail to falsify the theory yet, by definition, cannot prove it (Popper, 1968).

Identifying and clarifying existing theory also provides opportunities for framing a specific, causal, and generalizable research question. Such a question must be specific enough so that the planned research has a possibility of yielding a determinate answer. Drawing accurate causal inferences requires beginning with an explicitly causal question that explains observed variation rather than merely describing it. It also must be generalizable—that is, framed with

variables, their values, and relationships that could apply to a range of cases other than those studied.

We view causal questions as fitting into one of three possible modes,² defined by whether the researcher begins by specifying DVs, IVs, or both. One approach entails explaining a particular outcome or change in a specified DV, being initially agnostic about whether any particular factor was a major cause. Such research might focus on why international collaboration on some environmental problem—such as protecting the stratospheric ozone layer—is easier to achieve than collaboration on equally pressing problems—such as climate change, tropical deforestation, or biodiversity (Sebenius, 1991; Skolnikoff, 1990). A second approach investigates one or relatively few causal relationships by specifying both the IVs and DVs at the outset. This approach involves empirical testing of whether a specific IV has a hypothesized impact on a specific DV. Several large, collaborative research projects have investigated what impact, if any, rules of international environmental treaties have on observed environmental policies and behaviors (Jacobson & Brown Weiss, 1995; Victor, Raustiala, & Skolnikoff, 1997). A third approach analyzes the influences of a specified IV, as in Young's (in press) edited volume, which identifies a wide array of intended and unintended effects of international environmental regimes.

Innovative researchers also frame their concern about a particular environmental problem in broadly theoretical terms that clarify how the research results apply to other cases.³ For example, Princen's (1996) explanation of how the ivory trade ban established a norm against purchasing ivory would interest only those concerned about rhinos and elephants had he not framed it as an example of how bans establish ecological norms more effectively than other policies. The chapters in several recent edited volumes (Haas, Keohane, & Levy, 1993; Keohane & Levy, 1996; Young & Osherenko, 1993) showed how case studies can contribute to cumulative knowledge if cases that have inherent substantive appeal are selected and analyzed as examples of larger classes of cases.

2. As we note above, we seek to address those scholars who undertake case studies and qualitative research in the service of identifying and generalizing valid causal inferences, while recognizing the other, appropriate purposes to which case studies and qualitative research can be put.

3. Members of dissertation committees as well as discussants on panels at academic 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.

A tension often exists between specificity and generalizability. Stinchcombe (1968) distinguished between theoretical and empirical questions: The former involve larger, more conceptual variants of a question, with variables, values, and relationships framed in terms that could apply to cases besides the ones to be studied; the latter involve operational definitions of variables, values, and relationships that allow specific and falsifiable predictions about outcomes in the cases to be studied. Achieving the joint goals of internal and external validity requires that scholars continually reevaluate whether their work answers both a generalizable theoretical question and a matching and specific empirical question (Stinchcombe, 1968). For example, an analyst planning on using a case study of the North American Free Trade Agreement (NAFTA) to evaluate free trade's impact on the environment should frame the study's questions in generalizable terms that would make them equally useful for those studying the linkage of free trade (or its opposite) and environmental protection in the European Union, the World Trade Organization, or Mercosur.

DEVELOPING HYPOTHESES AND IDENTIFYING VARIABLES

Efforts to draw causal inferences require hypotheses formulated "so that they can be shown to be wrong as easily and quickly as possible," with the analyst identifying "what evidence would falsify" such hypotheses before research begins (King et al., 1994, p. 100). This requires identifying IVs, control variables (CVs), and DVs, their potential values, and their theorized causal relationships.⁴ Consider the claim that culture influences the likelihood that a particular state will comply with its international environmental commitments. The DV (compliance) and IV (culture) of the hypothesis are clear, as are potential values for the DV (low compliance or high compliance), but identifying the values of the IV proves more difficult. To make this hypothesis falsifiable, the researcher must categorize culture into at least two categories or values that she predicts correspond to higher and lower likelihoods of compliance. Whether categorizing culture as indigenous and nonindigenous, as strongly environmental, weakly environmental, and nonenvironmental, or by some other taxonomy, analysis of the impact of culture

4. George and Bennett (1997) argue 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.

on compliance requires delineation of the potential values of the variable culture.

Many researchers are interested in examining more than one causal relationship and, hence, more than one IV. To use case studies to draw valid inferences for such questions requires that the researcher include more cases than IVs in her analysis. Thus, the analyst faces a trade-off between the desire to draw valid causal inferences that aid our understanding of several causal relationships and the requirement that, to do so, requires analyzing more cases. In making this trade-off, we believe that drawing valid and unambiguous conclusions about a single causal claim contributes more to our knowledge of IEP than research that more boldly attempts to evaluate many causal claims but that uses an indeterminate research design with insufficient cases to validly substantiate any of the claims. The remedy involves ensuring the number of IVs is less than the number of cases—by studying more cases, by selecting cases to reduce the number of IVs that vary across the cases, or by refocusing “the study on the effects of particular explanatory variables . . . rather than on the causes of a particular set of effects” (King et al., 1994, p. 120).⁵

SELECTING CASES

Careful case selection grounded in existing theory lies at the heart of qualitative research that seeks to identify valid causal relationships. A particularly common threat to such design in IEP case studies arises from many researchers’ understandable and appropriate desire to make their analyses applicable to current policy debates. This desire produces an independent (but often inappropriate) pre-commitment to studying a particular case of current policy interest. Initiating research by selecting a case because of substantive interests and letting this case dictate the research questions often produces results with neither theoretical nor policy value. In most cases, theoretically “hot” questions cannot be answered using politically hot cases. For example, both scholars and practitioners have become increasingly concerned about regime effectiveness over the past decade. Yet, studies of “headline” issues—climate change, desertification, deforestation—will provide few, if any, useful insights into

5. Of course, the researcher should avoid reducing the number of IVs simply by ignoring other variables. Such an approach runs the risk of biasing the findings through omitted variable bias.

how to design international agreements to induce environmental improvement simply because the data needed to identify the sources of effectiveness in these cases are not yet available. Likewise, it may well be several years before the necessary data from the NAFTA experience will allow qualitative case studies to contribute meaningfully to the free trade and environment debate (Thompson & Strohm, 1996). Instead of analyzing hot issues, scholars concerned with causal inferences should select historical cases that will provide internally valid results but that also will, by intention, have characteristics that allow those results to be generalized accurately to the cases of policy concern.⁶

To understand how to select such cases first requires defining what constitutes a case. Cases are often defined in terms of environmental issues, such as biodiversity, whaling, or nuclear power in a specific country or region, or in terms of entities, such as an individual or organization. In contrast, following Eckstein (1975), we define a case as “a phenomenon for which we report and interpret only a single measure on any pertinent variable” (p. 85). The researcher delimits her cases by the unit of analysis she chooses. For example, a scholar interested in debt-for-nature swaps could conceive of her cases as individual debt-for-nature swaps drawn from a set of several swaps to be compared or as phases of a debt-for-nature swap, with different phases in the same swap being compared as the swap develops over time. Studies of a single environmental issue can and frequently do involve more than one case because they are based on a longitudinal analysis of the issue as a collection of several, embedded cases.

To isolate one variable’s influence from that of others, case study research selects cases so that the primary IV of interest varies but other IVs that might also influence the DV do not.⁷ For this reason, case selection is one of the most difficult and, in practice, poorly handled aspects of case study research. Drawing causal inferences requires comparing at least one case per value of the IV of interest. For ease of explication, 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. The minimum requirements for dem-

6. Descriptive analyses of existing 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 are usually unlikely to contribute to our understanding of causal mechanisms.

7. Quantitative research accomplishes this goal through large data sets and statistical algorithms.

onstrating that A caused B are showing (a) that, in one case, B is present when A is present; (b) that, in a similar counterfactual case (i.e., *ceteris paribus*), B is absent when A is absent; and (c) that A preceded B in time (Fearon, 1991). Evidence that A correlates with B says nothing about A's causal relationship with B unless, at a minimum, not-A also correlates with not-B.

Convincing causal case studies must exhibit variation in the value of the IV of interest and exhibit lack of variation in the value of other potentially explanatory IVs (i.e., CVs). Consider the popular proposition in the trade and environment literature—that trade liberalization causes convergence of environmental standards across countries. Because the transboundary character of some environmental problems (and not trade liberalization) may cause policy convergence, one should select cases in which the environmental problem is local—that is, not transboundary. If one observes convergence of initially disparate national policies for local environmental problems (especially if those problems are significantly different in magnitude) in situations characterized by liberalized trade and if one observes less convergence in nonliberalized settings, the proposition becomes more plausible because one was able to control a theoretically important CV. Analyzing a single environmental issue by dividing it into distinct time periods based on the values of the IV of interest or by conceptualizing the issue area as consisting of several subissue areas can provide easy mechanisms for observing change in one variable while other variables are held relatively constant.

Thus, for example, a researcher interested in the effectiveness of international regimes might compare levels of hazardous waste trade before and after establishment of the Basel Convention or evaluate how the number of fur seals killed varied before and after the adoption of the 1911 Fur Seal Treaty (Clapp, 1994; Mirovitskaya, Clark, & Purver, 1993). Such a design would not eliminate all alternative explanations of waste trade or fur seals killed but would eliminate many because, in each case, the nature of the environmental problem, and the number, character, and relative power of the actors involved, would have been held more constant than in a comparison across disparate cases.

Careful selection to limit variation in certain variables across the cases to be compared eliminates those variables as potential explanations of variation in a DV that would otherwise have confounded the analysis. This strategy can prove particularly valuable when an alternative explanatory variable, if it varied across the cases ana-

lyzed, would have theoretical priority. For example, Mitchell (1994) compared oil company compliance with two different but contemporaneous rules within a single international regime regulating oil pollution. By defining and selecting cases in this way, he ensured that economic influences (which would clearly have trumped regime influences) had the same impact on compliance with both rules (i.e., did not vary across the cases) and could, therefore, be excluded as explanations of variation in corporate compliance with the two rules.⁸

Unfortunately, the real world often fails to offer up cases that fit these demanding criteria. In such situations, careful counterfactual analysis can discipline the analyst to clarify the implications of the causal logic of propositions by imagining whether and how the world would have been different had the IV of interest had a different value (Fearon, 1991; Tetlock & Belkin, 1996, p. 3). For example, case studies on the effectiveness of international environmental regimes often ask questions such as, How would sulfur dioxide emissions in Western Europe have evolved if the regime on long-range transboundary air pollution had not been established? (Levy, 1995). Careful development of counterfactual scenarios can prompt the scholar to collect additional evidence that can strengthen conclusions drawn from the study. Recent attention to and development of criteria and methods for counterfactuals may help the analyst in avoiding the biases and wishful thinking that often plague such efforts (Dessler, 1991; Fearon, 1991; Tetlock & Belkin, 1996, pp. 16-37).

A researcher can increase a study's internal and external validity by seeking out "hard cases" in which the values of many of the CVs are "distinctly unfavorable" to the hypothesis being tested (Young, 1992, p. 165). As already noted, selecting cases so that variables other than the IV of interest have the same value excludes these variables as explanations of variation in the DV. In addition, however, if the researcher is analyzing only a few cases (and, therefore, cannot investigate cases in which CVs have different values), she can make more robust inferences by selecting cases in which the

8. Priority should also be given to ensuring that the cases selected exhibit variation in the value of the IVs of interest, "without regard to the values of the dependent variable" (King et al., 1994, p. 140). On one hand, because the analyst selects only a few cases from a larger population of cases, she must avoid intentionally picking cases because they support the hypothesized correlation of IV and DV that is allegedly being tested. On the other hand, by concerning herself only with the value of the IV, she must avoid the risk of accidentally selecting cases that exhibit no variation in the DV (King et al., 1994, pp. 142-146).

values of the CVs make it unlikely that the DV would have the value predicted by the IV. This approach reduces the likelihood of finding evidence of a hypothesized causal relationship but strengthens the power of the conclusion if such evidence is found (see, for example, Haas, 1990, p. 214; Mitchell, 1994; Young, 1989). Because even a single carefully conducted case study involves heavy resource demands, cases to be used in multiple case studies across issue areas should be selected so that each case leads either to identical or similar results (literal replication) or to different results for predicted reasons (theoretical replication) (Yin, 1994, p. 46).⁹

Selecting cases according to these methodological guidelines requires some initial knowledge of possible cases. One can often identify potentially valuable cases through discussion with other scholars, scanning the secondary literature, or, if necessary, conducting a pilot study. One may have to scrap cases already invested in, add cases not yet investigated, or completely reanalyze existing cases based on changing the unit of analysis. Often, the exigencies of the cases initially investigated push the scholar to shift the analysis from a broader analysis of variation across issue areas to a deeper analysis of variation across subcases within an issue area.

LINKING DATA TO PROPOSITIONS

With cases selected, the researcher can now evaluate the hypothesized causal relationship against the evidence. Having clearly defined all variables and their values (see above), the analyst must identify appropriate observable proxies or operationalizations (Bernauer, 1995; Dessler, 1991, p. 339; Levy, Young, & Zürn, 1995; Young, 1992). In translating general theoretical statements into more specific empirical statements (Stinchcombe, 1968), the variables must be defined and operationalized so that data relate to the theoretical construct as accurately as possible (construct validity). General theo-

9. In literal replication, the researcher chooses two cases that exhibit the same value on the DV. If proposed IVs turn out to have the same values and all other variables can be held constant, the model becomes more plausible. This approach follows the logic of repeated experiments under identical conditions. In theoretical replication, the researcher selects cases whose IVs (or, less preferably, DVs) exhibit different values. If different values of the IV are associated with identical values of the DV, the proposition is probably wrong. If, for theoretically predicted reasons (different values of the IV), the outcomes vary, one has carried out a theoretical replication, comparable to a repeated experiment under different, but controlled, conditions.

ries, precisely because they seek to explain outcomes across a wide array of cases, often define variables in nonspecific terms. For example, most scholars accept that interests influence behavior in international relations. However, constructing falsifiable predictions for a particular case requires defining interests in such a way that one type of interest (i.e., a value for the variable interests) predicts a behavior different than that predicted by a different type of interest.

Thus, Sprinz and Vaahtoranta (1994) conceptualized interests in international environmental policy as consisting of a state's "environmental vulnerability" and "abatement costs." In the case of stratospheric ozone depletion, they operationalized these two variables as the rate of skin cancer and the intensity of chlorofluorocarbon (CFC) consumption (p. 89). Although these proxies do not capture the full conceptual richness of the variables ecological vulnerability or abatement costs, they correspond reasonably well to these concepts and have the virtue of being readily observable. Appropriate, reliable, and observable indicators of complex conceptual variables often prove difficult to find (Homer-Dixon, 1996). Given these problems of operationalizing variables, the in-depth analysis permitted by case study techniques offers the potential to observe a variety of indicators for each conceptual variable, thereby more accurately and reliably approximating the values of the conceptual variables of interest. Developing such multiple proxies to triangulate on a single value of the conceptual variable increases the researcher's confidence in her assessment even while recognizing that any one of these proxies may incompletely capture the value of the variable concerned. For example, Levy (1995) evaluated various indicators to determine both the extent and the causes of European states conforming their policies to the dictates of the European acid precipitation regime.

Examining different proxies for the same theoretical construct may not only strengthen construct validity and reliability but also lead to new theoretical insights. In a project on the implementation of international environmental commitments, Victor et al. (1997) analyzed various indicators of implementation. They found that formal legal compliance with environmental treaty commitments is often high, whereas behavioral effectiveness measuring whether a regime shapes actor behavior is rather low. Informal commitments, on the other hand, tend to be associated with a lower degree of compliance but with higher behavioral effectiveness. This finding

opened an avenue for innovative explanations of variation across different indicators for effectiveness.

Another important advantage of qualitative causal case studies is their ability to bring to light variables that were not initially thought to play an important causal role and hence were not included in the original research design. We believe that it is possible to maintain a healthy tension between the structure imposed by delineating IVs, DVs, CVs, and their relationships before the study begins and an open-mindedness that the full story may not be captured within the bounds established by this initial structure. For example, the importance of capacity as a factor conditioning the effectiveness of international environmental protection was an inductively derived insight from (rather than a deductively derived hypothesis tested by) the case studies of several environmental regimes in Haas et al.'s (1993) book.

Armed with clear definitions and operationalizations of the variables, the researcher can proceed to gather data on the value of each IV, CV, and DV for each case. Data collection can include reviews of primary and secondary literature and documents, structured or open interviews and surveys, direct or participant observation, or collecting quantitative data. For example, in examining whether a particular international agreement contributed to the desired pollution reductions, a researcher might collect both systematic and anecdotal evidence from various sources regarding changes in pollution levels over time, changes in pollution control technologies and economic trends that might produce exogenous changes in pollution levels, and decision makers' and activists' views on whether and why pollution levels changed over time and how the agreement contributed to these reductions.

EXAMINING CORRELATIONS AND CAUSAL PATHWAYS

Unlike quantitative researchers, qualitative case study researchers cannot turn to well-established and well-accepted procedures for analyzing the data they collect. We believe, however, that good causal analysis can be promoted by simple and systematic comparisons of predicted and observed values of the DV, evaluation of causal narratives, and evaluation of rival hypotheses. Predicting the values of the DV for each case based on the theory being tested and the observed values of the IVs and CVs in each case provides a structure for evaluating whether the empirical evidence conforms

with theoretical expectations.¹⁰ Failure of the predicted DV values to match the observed DV values indicates that the theory does not generalize to the case in question, that the theory deals with a necessary but not sufficient condition for the DV to assume the predicted value, or that the theory lacks explanatory power more generally.¹¹

We have found that a simple table summarizing the values of the variables in each case imposes beneficial rigor on causal case analyses. Each row corresponds to a case, with columns established for the values of each IV, the values of each CV, the predicted value for the DV (based on the values of the IVs and CVs), and the observed value of the DV. Such a table forces the researcher to be explicit about what theory predicts in the cases at hand and to focus on the degree to which the IV and DV covary as predicted. If the IVs of interest and DV correlate, then, as in regression analysis, the analyst needs to determine whether the covariation is evidence of a causal relationship or simply spurious covariation caused by other variables. This step may involve including the values of other CVs in the chart to check for their covariation with the DV. If case selection succeeded at holding the CVs constant, one can eliminate rival claims that those variables caused the observed variation in the DV.¹² Often, however, CVs that originally appeared to be constant across cases prove, on deeper research, to actually have varied. If the CVs were not actually held constant, the researcher should reconsider her case selection, turn CVs into explanatory variables, and make a particularly intensive effort at investigating rival propositions (see below).

Consider a study evaluating how scientific consensus influences the outcome of international environmental negotiations. The IV might be defined as level of scientific consensus on the causes of an environmental problem, with potential values of high or low; the DV

10. Often, this step requires that the analyst temporarily ignore already-obtained information about the value of the DV: Properly evaluating the theorized relationship requires comparing the observed value of the DV with the value of the DV as predicted by theory. Despite already knowing the actual value of the DV in one or more of the cases being studied, the researcher should nonetheless identify what the theory under investigation would predict about that DV's value given the values of the IVs and CVs for that case.

11. 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.

12. CVs that have a value that theory predicts would cause the DV to assume a value different than the observed value can also be eliminated as plausible explanations of variation in the DV.

might be defined as negotiation of international environmental agreement, with potential values of negotiated successfully or not negotiated successfully. Evaluating the internal validity of this hypothesis requires comparing one case exhibiting high scientific consensus regarding the causes of an environmental problem with another case exhibiting low consensus. Four types of results might arise: (a) a negotiated agreement in the first case but not the second, (b) a negotiated agreement in the second but not the first, (c) negotiated agreements in both cases, and (d) no negotiated agreement in either case. In results (c) and (d), no causal relationship appears to exist between the IV and DV. Result (a) provides initial support that scientific consensus may facilitate negotiation of an agreement, whereas result (b) provides initial support that scientific consensus may inhibit negotiation of an agreement.

If the researcher finds covariation between the IV of interest and the DV (with CVs held constant), she should go on to provide a plausible causal narrative of why and how the IV caused variation in the DV.¹³ Case studies have a major advantage over quantitative methods in this regard, because they allow disaggregated and in-depth analysis of such "causal mechanisms" or "causal pathways" (Dessler, 1991, p. 352). Detailed causal narratives or "process tracing" (George, 1979) are more than mere storytelling. Essentially, examining the causal pathways that link an IV and DV in a particular case allows the researcher to show how and why these two variables covary rather than merely showing that they covary. Several recent research projects have used this method to evaluate the different causal pathways by which regimes wielded influence (Bedarff, Bernauer, Jakobeit, & List, 1995; Haas et al., 1993; Keohane & Levy, 1996; Victor et al., 1997; Young, 1996).

Process tracing involves breaking down an overarching causal relationship into a set of smaller causal links in a larger causal chain. For example, demonstrating the causal pathway by which scarcity of natural resources due to environmental degradation causes acute conflict requires demonstrating that the environmental degradation caused migration or exacerbated interethnic rivalries between mi-

gratory groups and indigenous populations in ways that could have induced the observed conflict (Homer-Dixon, 1991, pp. 85-86, 1996).

The internal validity of the study can be further enhanced if the researcher explicitly considers alternative explanations and finds them "to be less consistent with the data and/or less supportable by available generalizations" (George, 1979, pp. 57-58). Standard theoretical explanations for variation in the DV should provide a list of "likely suspects" for these rival hypotheses. Careful case selection should have eliminated many of these potentially explanatory variables by holding them constant. In the remaining cases, the researcher should evaluate whether evidence of the influence of these rival variables provides 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, especially when strong a priori reasons exist to accept the power of these explanations. For example, a researcher interested in whether a nongovernmental organization's (NGO's) activism caused a change in a state's negotiating position may find competing evidence that a change in economic interests caused the negotiating position change. If so, she should accept the latter rival explanation simply because of the strong theoretical and empirical evidence across many areas of international relations that short-term economic interests are a powerful determinant of state negotiating positions and behavior. The strategy of explicitly evaluating rival hypotheses constitutes the most effective defense against charges that the outcome observed in the cases was caused by variables other than those claimed by the researcher.

Two examples illustrate this procedure. One explanation of why governments negotiated and signed the Montreal Protocol in 1987 contends that scientific consensus, stimulated mainly by the discovery of the ozone hole, produced the proximate pressure for agreement (Haas, 1992a; Parson & Greene, 1995). Testing this causal claim requires the analyst to refute the rival hypothesis that governments completed the agreement in response to Dupont—the world's leading producer of CFCs—(and subsequently other producers) removing their objections to the agreement for economic, not scientific or environmental, reasons. The analyst would thus need to demonstrate whether and why this rival explanation is empirically inaccurate or incomplete in explaining the observed timing of the agreement. Similarly, Bernauer and Moser (1996) sought to evaluate the

13. As George (1979) notes, "to assess whether a statistical correlation between independent variables and the dependent variable is of causal significance, the investigator subjects a single case 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" (p. 46).

hypothesis that the regime for protecting the river Rhine had led to decreased pollution levels. They developed several plausible causal narratives leading from the presumed cause to the observed outcome. Process tracing via open interviews with decision makers as well as analysis of pollution and regulatory data and the secondary literature showed that transboundary regulation probably had caused only a fraction of the observed pollution reduction. The interviews in particular shed light on two more plausible rival explanations of pollution reductions: a secular decline in heavy industries and coal mining along the Rhine, and a general greening of domestic politics. Subsequent analysis showed that these two variables had not stayed constant over time and, indeed, appeared to correlate with pollution reductions, making them likely candidates for a causal explanation. By carefully looking for and evaluating these rival hypotheses, the analysts avoided inaccurately attributing causal power for the pollution reductions to the international regime.

The foregoing discussion highlights a unique risk faced by any scholar attempting qualitative case study work. Case studies provide no fallback position if the IV of interest proves not to cause variation in the DV. The methodological need to exclude cases that exhibit variation in explanatory variables other than the IV runs the risk that if the IV of interest does not explain variation in the DV, the researcher will have to (a) examine new cases to see if, under other conditions, the hypothesized relationship does exist or (b) collect new data from the cases studied to determine whether other variables previously considered noncausal actually caused the observed variation in the DV.¹⁴ Many scholars can recount studies involving carefully selected cases 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

As a final step, researchers undertaking causal qualitative case studies should evaluate the extent to which the propositions sup-

14. Quantitative research, by contrast, statistically manipulates data on several potentially causal IVs to control for the variation of some variables when investigating the causal relationship of primary interest. If the existence of this primary relationship is refuted, the quantitative researcher can transform IVs into CVs, often using the same data set. In qualitative research, careful case design and selection can mitigate these risks but not eliminate them.

ported by the case evidence are relevant to other cases.¹⁵ Assessing the study's external validity—that is, whether the causal relationships present in the cases examined also operate in a larger population of cases—links the findings back to broader theoretical debates. If the cases studied have conditions that differ significantly from those of the larger population of cases, accurate generalization becomes far more difficult. Convincing policy makers to adopt (or avoid) a particular policy (or scholars to accept a particular theory) requires demonstrating that the conditions in the case studied look sufficiently like the conditions in the targeted policy area to support an expectation that the same causal influences will operate there.

It might be argued that case studies provide a weaker foundation for generalization than quantitative analyses. In a quantitative, large-*N* study, achieving internal validity simultaneously achieves significant external validity because internally valid findings are at least relevant to the large set of cases studied. Equally important, high-quality quantitative analysis can identify both the average correlation of the IV with the DV over a range of values for the CVs and also any interactive effects between the IV and various CVs. In contrast, the design of most causal case studies implies that the correlation of the IV with the DV is known only for a single value of the CV—that is, the value of the CV for which the cases were held constant. Consider, for example, a quantitative and a qualitative study of how the activism of environmental NGOs influences regime formation. Imagine that the only variable that theory suggests needs to be controlled for is the existence of a hegemon.¹⁶ Assuming (unrealistically) that the world presented sufficient cases, the quantitative analyst, armed with data on cases with and without NGO activism, with and without hegemons, and with and without regimes not only could identify the average impact NGOs had on regime formation but also might find that this average 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 if allowed to vary across the cases, would have selected cases to hold hegemony con-

15. Indeed, many scholars in the constructivist tradition argue that the importance of contextual contingency demands a healthy skepticism about even limited generalizations.

16. Haas (1990, 1992b) has developed an extensive theoretical model of how epistemic communities contribute to regime formation. Young (1983) has developed models distinguishing imposed from negotiated and spontaneous regimes based on the role played by hegemons.

stant, perhaps selecting only cases in which hegemony was absent. If properly conducted, the qualitative study should find the same large influence of NGOs on regime formation as the quantitative study found among cases in which hegemony was absent. A qualitative study also would have developed a clearer sense of exactly how the NGOs fostered regime formation. However, the qualitative analyst might be tempted to make the general claim that NGOs play a major role in regime formation when, in fact, that claim is true only for cases in which hegemony is absent.

Although quantitative methods may have many advantages in principle, in practice, these advantages often prove elusive, chiefly because many theoretically and practically interesting concepts in IEP have proved impossible to quantify in a meaningful way. As a result, qualitative case study work often has proved superior in generating generalizable causal arguments in IEP. In any event, the example just given illustrates the relationship of internal and external validity inherent in qualitative case studies. Internal validity is a precondition for external validity. Selecting cases to hold the value of certain variables constant increases the internal validity of causal inferences derived from the study but simultaneously limits the range of cases to which one can validly generalize.

Analysts can mitigate this problem in three ways. First, of course, they can do so by conducting additional case studies in which cases are selected so that the CV has a different value. Second, the analyst can increase generalizability by selecting hard cases in which, in all the cases analyzed, the CV has a value at which the findings from the case studies will be most generalizable. Thus, in the example above, because the existence of a hegemon is alleged to facilitate regime formation, cases selected to evaluate whether NGOs actually facilitate regime formation would be best chosen so that a hegemon is not in evidence. By doing so, the analyst who finds support for the influence of NGOs can make the cautious but appropriate claim that these cases support the contention that NGOs facilitate regime formation even when no hegemon is available to support this effort and that, therefore, one should expect NGOs to contribute to regime formation even when a hegemon is available. Third, the analyst can design the case studies to build on past case studies and contribute to a collaborative research program. By literal or theoretical replication of other cases, choosing cases to evaluate how different values of CVs influence the causal relationship between the IV of interest and the DV, the researcher can contribute to collaborative research

that collectively establishes generalizable results through joint analysis of a large set of cases (Yin, 1994).

The researcher should remain cautious about generalizing too broadly. Consider, for example, an analysis that convincingly demonstrated that the collective scientific research fostered by the long-range transboundary air pollution regime caused adoption of policies reducing SO₂ and NO_x emissions in Europe (Levy, 1993). A generalization such as "collective scientific research programs that identify environmental harms will induce states to adopt policies to redress those environmental harms" might apply only in the highly interdependent European context and not in the less integrated and less developed African, Asian, or Latin American contexts. Assessing external validity requires determining whether the values at which CVs were held constant were crucial to the observed influence of the IV of interest on the DV. The case study findings may still be generalizable so long as the values of the CVs are common to most cases.

Conclusion

Causal case studies are a widely used method of research in IEP. Yet, causal case study methodology remains underdeveloped and rarely appears on political science teaching curricula. In carrying out qualitative case studies, researchers face myriad potential pitfalls. The ease with which one can identify poorly constructed case studies in the IEP literature and elsewhere in the international relations literature testifies to the difficulty of designing and conducting them well. We regard this article as part of a broader effort to remedy this deficiency.¹⁷ Much of our knowledge of IEP is likely to be drawn from qualitative case studies—the more so because qualitative case studies have the distinct advantage over quantitative studies of helping us to understand in more detail the nature of causal relationships (George & Bennett, 1997; King et al., 1994, p. 86).

We have outlined six key steps that we believe constitute a solid foundation for rigorous, qualitative, and causal case-study research on IEP. Although recognizing the important contributions to IEP research of constructivist, interpretive, descriptive, and other case

17. See George and Bennett's (1997) volume that addresses case study methodology in depth.

study approaches, we have sought to delineate a set of guidelines for those scholars undertaking positivist, causal, variation-finding case studies. In this vein, we believe that good causal research begins by identifying an important and compelling research question. This question is then transformed into testable hypotheses by clearly delineating variables and their potential values. Equipped with these resources, we have argued that case selection provides a strong and crucial, but all too often ignored, foundation for validly identifying causal relationships and generalizing those findings to other cases. We note the hazards of selecting cases before developing a clear-cut theory and also emphasize that theoretically hot questions can often not be answered with politically hot cases. Case study research requires explicit attention to existing theory so as to use identification of IVs and CVs to predict the DV. By using a simple table to examine the degree to which observed variation in the DV corresponds to variation predicted by theory and by using process tracing of the pathways by which those factors caused that variation, good case studies can deepen our understanding of the complicated social and political processes at work in international environmental politics. A researcher can and should conclude her research by clearly identifying the cases to which the findings can (and cannot) be appropriately generalized. Even though specificity and generalizability may need to be traded off, ensuring that findings are internally valid must take precedence over claiming they are externally valid.

Conclusions drawn from qualitative case studies are often difficult to generalize, and such research can test only the causal relationships between a relatively few IVs and one DV at a time. Because of these problems, those undertaking qualitative case studies will be best served by following Mark Twain's (1894) dictum to "put all your eggs in the one basket and—WATCH THAT BASKET" (p. 15). The procedures outlined above are designed to reduce, if not eliminate, the risks that are inherent in qualitative empirical approaches. We believe that researchers attentive to these six steps can improve the analytic rigor of their work and, thereby, contribute to the growing scholarly 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.

Manuscript submitted May 22, 1997; revised manuscript accepted for publication October 12, 1997.

Acknowledgments

The authors would like to express their appreciation to Detlef Sprinz as well as to Peter Haas, Harold Jacobson, Peter Moser, Roy Suter, Arild Underdal, and two anonymous reviewers for extremely helpful comments on earlier drafts of this article.

References

- Baldwin, D. A. (1979). Power analysis and world politics: New trends versus old tendencies. *World Politics, 31*(2), 161-193.
- Bedarff, H., Bernauer, T., Jakobeit, C., & List, M. (1995). Transferzahlungen in der internationalen umweltpolitik: Das problem der transaktionskosten beim tausch von finanzhilfe gegen umweltschutzmassnahmen [Transfer payments in international environmental policy: The problem of transaction costs in exchanges of financial aid for environmental protection measures]. *Zeitschrift für Internationale Beziehungen, 2*(2), 317-345.
- Bernauer, T. (1995). The effect of international environmental institutions: How we might learn more. *International Organization, 49*(2), 351-377.
- Bernauer, T., & Moser, P. (1996). Reducing pollution of the Rhine River: The influence of international cooperation. *Journal of Environment & Development, 5*(4), 391-417.
- Breitmeier, H. (1996). *Wie entstehen globale umweltregime?: Der konfliktstrag zum schutz der ozonschicht und des globalen klimas* [How do global environmental regimes emerge? Conflicts in protecting the ozone layer and global climate]. Opladen: Leske+Budrich.
- Clapp, J. (1994). The toxic waste trade with less-industrialized countries: Economic linkages and political alliances. *Third World Quarterly, 15*(3), 505-518.
- Dessler, D. (1991). Beyond correlations: Toward a causal theory of war. *International Studies Quarterly, 35*(3), 337-355.
- Eckstein, H. (1975). Case study and theory in political science. In F. Greenstein & N. Polsby (Eds.), *Handbook of political science: Vol. 7. Strategies of inquiry* (pp. 79-137). Reading, MA: Addison-Wesley.
- Fearon, J. D. (1991). Counterfactuals and hypothesis testing in political science. *World Politics, 43*(2), 169-195.
- George, A. L. (1979). Case studies and theory development: The method of structured, focused comparison. In P. Lauren (Ed.), *Diplomacy: New approaches in history, theory, and policy* (pp. 43-68). New York: Free Press.
- George, A., & Bennett, A. (1997). *Case studies and theory development*. Manuscript in preparation.
- George, A., & Bennett, A. (1997, March). *Developing and using typological theories in case study research*. Paper presented at the International Studies Association Conference, Toronto.
- Haas, P. M. (1990). *Saving the Mediterranean: The politics of international environmental cooperation*. New York: Columbia University Press.
- Haas, P. M. (1992a). Banning chlorofluorocarbons. *International Organization, 46*(1), 187-224.
- Haas, P. M. (1992b). Epistemic communities and international policy coordination. *International Organization, 46*(1), 1-35.

- Haas, P. M., Keohane, R. O., & Levy, M. A. (Eds.). (1993). *Institutions for the earth: Sources of effective international environmental protection*. Cambridge, MA: MIT Press.
- Homer-Dixon, T. (1991). On the threshold: Environmental changes as causes of acute conflict. *International Security*, 16(2), 76-116.
- Homer-Dixon, T. (1996). Strategies for studying causation in complex ecological political systems. *Journal of Environment & Development*, 5(2), 132-148.
- Jacobson, H. K., & Brown Weiss, E. (1995). Improving compliance with international environmental accords. *Global Governance*, 1(2), 119-148.
- Keohane, R. O., & Levy, M. A. (Eds.). (1996). *Institutions for environmental aid: Pitfalls and promise*. Cambridge, MA: MIT Press.
- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
- Levy, M. (1993). European acid rain: The power of tote-board diplomacy. In P. Haas, R. O. Keohane, & M. Levy (Eds.), *Institutions for the earth: Sources of effective international environmental protection* (pp. 75-132). Cambridge, MA: MIT Press.
- Levy, M. A. (1995). International cooperation to combat acid rain. In Fridtjof Nansen Institute (Ed.), *Green globe yearbook: An independent publication on environment and development* (pp. 59-68). Oxford, UK: Oxford University Press.
- Levy, M. A., Young, O. R., & Zürn, M. (1995). The study of international regimes. *European Journal of International Relations*, 1(3), 267-330.
- Mirovitskaya, N., Clark, M., & Purver, R. G. (1993). North Pacific fur seals: Regime formation as a means of resolving conflict. In O. R. Young & G. Osherenko (Eds.), *Polar politics* (pp. 22-55). Ithaca, NY: Cornell University Press.
- Mitchell, R. B. (1994). *Intentional oil pollution at sea: Environmental policy and treaty compliance*. Cambridge, MA: MIT Press.
- Moon, J. D. (1975). The logic of political inquiry: A synthesis of opposed perspectives. In F. Greenstein & N. Polsby (Eds.), *Handbook of political science* (Vol. 1, pp. 131-207). Reading, MA: Addison-Wesley.
- Parson, E. A., & Greene, O. (1995). The complex chemistry of the international ozone agreements. *Environment*, 37(2), 16-22.
- 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), 147-176.
- Princen, T., Finger, M., & Manno, J. (1995). Nongovernmental organizations in world environmental politics. *International Environmental Affairs*, 7(1), 42-58.
- Ragin, C. C. (1994). *Constructing social research: The unity and diversity of method*. Thousand Oaks, CA: Pine Forge.
- Sebenius, J. K. (1991). Designing negotiations toward a new regime: The case of global warming. *International Security*, 15(4), 110-148.
- Skolnikoff, E. B. (1990, Summer). The policy gridlock on global warming. *Foreign Policy*, pp. 77-93.
- Sprinz, D., & Vaahoranta, T. (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.
- Tetlock, P. E., & Belkin, A. (Eds.). (1996). *Counterfactual thought experiments in world politics: Logical, methodological, and psychological perspectives*. Princeton, NJ: Princeton University Press.
- Thompson, P., & Strohm, L. A. (1996). Trade and environmental quality: A review of the evidence. *Journal of Environment & Development*, 5(4), 363-388.
- Twain, M. (1894). *The tragedy of Pudd'nhead Wilson; And, the comedy, those extraordinary twins*. Hartford, CT: American.
- Victor, D., Raustiala, K., & Skolnikoff, E. (1997). *The implementation and effectiveness of international environmental commitments*. Cambridge, MA: MIT Press.
- Wapner, P. (1995). Politics beyond the state: Environmental activism and world civic politics. *World Politics*, 47(3), 311-340.
- Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Young, O. R. (1983). Regime dynamics: The rise and fall of international regimes. In S. D. Krasner (Ed.), *International regimes* (pp. 93-113). Ithaca, NY: Cornell University Press.
- Young, O. R. (1989). *International cooperation: Building regimes for natural resources and the environment*. Ithaca, NY: Cornell University Press.
- Young, O. R. (1992). The effectiveness of international institutions: Hard cases and critical variables. In J. N. Rosenau and E.-O. Czempiel (Eds.), *Governance without government: Change and order in world politics* (pp. 160-194). New York: Cambridge University Press.
- Young, O. R. (1994). *International governance: Protecting the environment in a stateless society*. Ithaca, NY: Cornell University Press.
- Young, O. R. (1996). *Regime effectiveness: Taking stock*. Unpublished manuscript, Dartmouth College.
- Young, O. R. (Ed.). (in press). *Regime effectiveness*. Cambridge, MA: MIT Press.
- Young, O. R., & Osherenko, G. (Eds.). (1993). *Polar politics: Creating international environmental regimes*. Ithaca, NY: Cornell University Press.

Ronald Mitchell is an assistant professor in the Department of Political Science at the University of Oregon. He earned his PhD in public policy at Harvard University. His book *Intentional Oil Pollution at Sea: Environmental Policy and Treaty Compliance* (MIT Press, 1994) received the Harold and Margaret Sprout Award for 1995 from the International Studies Association for the best book on international environmental issues. He has published articles in *International Organization*, *Journal of Theoretical Politics*, *Nonproliferation Review*, and *Environment*.

Thomas Bernauer is an assistant professor of international relations at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland. From 1988 to 1992, he was a research associate at the United Nations Institute for Disarmament Research (UNIDIR) in Geneva, and from 1992 to 1994, he was a research fellow at Harvard University. His recent publications have appeared in *Security Dialogue*, *Environmental Politics*, *International Organization*, *Zeitschrift für Internationale Beziehungen*, and *Journal of Environment & Development*. He is coauthor, with Dieter Ruloff, of *Soliciting Cooperation From Critical States: The Politics of Positive Incentives in Arms Control* (University of South Carolina Press, forthcoming).