The Policy Studies Journal, Vol. 35, No. 2, 2007

Causal Assessment in Small-N Policy Studies

Paul F. Steinberg

The identification of cause-and-effect relationships plays an indispensable role in policy research, both for applied problem solving and for building theories of policy processes. Historical process tracing has emerged as a promising method for revealing causal mechanisms at a level of precision unattainable through statistical techniques. Yet historical analyses often produce dauntingly complex causal explanations, with numerous factors emerging as necessary but insufficient causes of an outcome. This article describes an approach that renders complex causal narratives more analytically tractable by establishing measurement criteria for ranking the relative importance of component causes. By focusing on subjectively useful measurement attributes, the approach is well suited to the policy sciences' unique combination of explicitly normative aspirations and a commitment to the systematic assessment of causal claims.

KEY WORDS: causation, causal importance, process tracing, policy analysis, qualitative research methods

Central to the aims of public policies, and the political constituencies supporting them, is the hope of having a causal impact on some aspect of the world. It is hoped that welfare-to-work programs will lead to a decline in chronic unemployment; that the international whaling regime will cause threatened species to rebound; and that health education campaigns will reduce HIV transmission. As Pressman and Wildavsky (1973, p. xxi) observed, "Policies imply theories. Whether stated explicitly or not, policies point to a chain of causation between initial conditions and future consequences. If X, then Y." Accordingly, while causal theories play a role in many areas of social inquiry, they are vital to the practice of policy analysis, where they are used to diagnose problems, project future impacts of new regulations, and evaluate the effectiveness of—and assign responsibility for—past interventions (Chen, 1990; Lin, 1998; Young, 1999). Causal assessment plays an equally important role in the policy process tradition, as researchers identify the causal factors shaping policy agendas, decision-making styles, state–society relations, and the dynamics of stability and change (Baumgartner & Jones, 1993; Rochon & Mazmanian, 1993; Sabatier, 1999).

Against this backdrop, this article focuses on an issue of special importance to policy-oriented political scientists: how to assess causal impacts in small-N research settings. This question is motivated by a scenario all too familiar to policy research-

0190-292X
Published by Blackwell Publishing. Inc., 350 Main Street, Malden, MA 02148, USA, and 9600 Garsington Road, Oxford, OX4 2DQ.

ers. Having undertaken a thorough historical analysis to identify the causal processes leading to an outcome of interest, the investigator discovers that the outcome resulted from a multifaceted stream of events, with numerous variables emerging as necessary but insufficient to produce the outcome. Is it possible in these scenarios to make meaningful statements regarding the relative causal importance of the component factors? After all, a great attraction of regression analysis is its precise delineation of the relative contributions of diverse independent variables, reported in the form of partial correlation coefficients. How might relative causal importance be operationalized in the context of small-N studies, that is, when the cases number fewer than a dozen and often as few as one to three?

To render this question more tangible, let us consider a case of successful rural development that has recently attracted considerable attention from international donor agencies. The village of Ait Iktel, in the High Atlas Mountains of southern Morocco, has over the past decade achieved dramatic improvements in public education and water and power services. According to World Bank analyses (Mernissi, 1997; World Bank, 2003, pp. 75–76), this outcome resulted from the activities of a local nongovernmental organization (NGO) that mobilized community interest, building on the village's considerable social capital, and drew support from emigrants working abroad and from the French and Japanese governments. This NGO arose in the context of macro-structural shifts—including political liberalization by King Hassan II in the mid-1990s, which facilitated growth in advocacy-oriented NGOs, and tentative moves toward decentralization of state-sponsored education. Furthermore, the success of Ait Iktel relative to other villages with similar levels of social capital, and which experienced identical structural shifts, was largely because of two dynamic local leaders who had strong capabilities in external fundraising and project management. In this case a variety of conditions—political liberalization, decentralization, external resources, social capital, and a local NGO under entrepreneurial leadership—were all probably necessary to produce the observed outcome of improved social services.[1]

The Ait Iktel case is not unusual in this regard. A similar result is found in the report of the 9/11 Commission, the U.S. congressional body charged with assessing the conditions that made the September 11 attacks possible (Kean & Hamilton, 2004). The commission report illustrates both the indispensability of small-N research for policy analysis and the dauntingly complex explanations that can result. Table 1 provides a sample of the dozens of factors identified by the commission as significantly contributing to American vulnerability, including inadequate information sharing among agencies, sporadic congressional attention, foregone opportunities to attack Bin Laden in the late 1990s, the ease of producing false documentation, the Federal Bureau of Investigation (FBI)'s emphasis on legal prosecution rather than long-term investigations, airliner resistance to burdensome security measures, the absence of federal air marshals, and mayors' resistance to law enforcement training by the Immigration and Naturalization Service.

With their multifaceted and often nested sets of causal conditions, the Ait Iktel and 9/11 examples are typical of the phenomena studied by policy researchers. Yet the question of how to adjudicate the relative importance of component causes in

Table 1. Some Causes of American Vulnerability to the 9/11 Attacks

Inadequate information sharing among agencies.
FBI emphasized legal cases rather than long-term investigations.
Outdated INS technology produced incomplete terrorist watch list.
Ease of false documentation.
Absence of federal air marshals.
Some mayors resisted INS training of local law enforcement.
FAA had poorly developed intelligence functions.
Suicide attacks were not perceived as the primary threat to civil aviation.
Airliners resisted burdensome security measures.
National Security Agency focused exclusively on foreign intelligence.
Department of Defense was not informed that a second plane had been hijacked.
Congressional attention to terrorism was "episodic and splintered."
Planned attacks against Bin Laden in 1998 and 1999 were canceled.
Afghan tribal leaders failed to act on a U.S. offer to pay for the capture or death of Bin Laden.

Source: Kean and Hamilton (2004).
FBI, Federal Bureau of Investigation; INS, Immigration and Naturalization Service; FAA, Federal Aviation Administration.

small-N settings has received surprisingly little attention in the research literature, as I detail below. The omission is a significant one. When causal assessment is undertaken for applied purposes, agents of change have limited resources and need to know how to prioritize their efforts, whether they are trying to win an election, influence a policy process, or improve a social outcome. Determinations of relative causal importance are also indispensable for theory building, which requires evaluations of the relative explanatory power of competing causal hypotheses. Finally, this problem of multiple necessary conditions—described by Fischer (1970, pp. 175–77) as indiscriminate pluralism—carries broader implications for the role of qualitative methods in policy research and in social science inquiry generally. The great promise of small-N methods, and of historical process tracing in particular, is the richness, complexity, and nuance that they provide for constructing valid causal explanations (George & Bennett, 2005; George & McKeown, 1985). To be of value, however, process tracing must not only help us to reveal complexity, but to make sense of it. Failure to do so would call into question its value as a tool for causal assessment.

The goal of this article is to provide a number of approaches for ranking causal importance in small-N settings that are justifiable on methodological grounds and useful for policy purposes. My argument rests on three basic points. First, to tackle the problem of indiscriminate pluralism we must recognize that causal outcomes are the result of sets of conditions. The most widely understood definition of causation, and the one used in this article, holds that a factor is a cause if its presence increases the likelihood of an outcome (Gerring, 2005). Yet as Roberts (1996, p. 75) explains, "one event does not cause another; a match carelessly thrown away does not cause a fire. Rather, a set of conditions—a thrown away match, dry twigs, and the presence of oxygen—causes the fire." In other words, a cause-and-effect relationship is an emergent property of a set of interacting conditions. The challenge, in causal analysis generally and in this article, is to say something meaningful about isolated components—which is necessary to make the world intelligible yet must be accomplished in a world that is in fact highly interconnected.

Second, the way out of this apparent dilemma is to tether the analysis to a pragmatic analytic goal and associated measurement attributes that provide a basis for discerning among component causes. Following Lazarsfeld and Barton's (1951) insight, later popularized by Sartori (1970), the identification of subjectively useful attributes is a prerequisite for objective measurement. "[B]efore we can rank objects or measure them in terms of some variable, we must form the concept of that variable. Looking at the material before us in all its richness of sense-data, we must decide what attributes of the concrete items we wish to observe and measure: do we want to study 'this-ness' or 'that-ness' . . . ?" (Lazarsfeld & Barton, p. 155). I argue that this is true not only of descriptive rankings, but of causal measurement as well. The implications of this approach to causal assessment run far deeper than the commonsense notion that a researcher must privilege some aspects of a causal explanation over others according to the purpose of the study. Instead this cuts to the very meaning and operationalization of causation: Intuitively appealing terms like causal strength and causal impacts only assume meaning when tied to specific analytic goals that serve as axes along which one can array component causes.

Third, in the context of policy research it is helpful to think in terms of three analytic goals in particular: covariance, leverage, and attribution. *Covariance* is the metric underlying approaches such as regression analysis or controlled comparison of cases, which emphasize the degree to which changes in antecedents (independent variables) are associated with changes in outcomes (dependent variables). Covariance analysis copes with complexity by reducing the causal question to a simpler form, explaining only the marginal difference in outcomes across cases rather than the fuller origins of these outcomes, and removing from the analysis those antecedents that may be important causes of the outcomes but are constant across all cases. I argue that while covariance is a powerful device for overcoming indiscriminate pluralism, it carries a number of drawbacks and does not exhaust the possibilities for addressing the problem. In contrast, *leverage* focuses on those antecedents most susceptible to manipulation and arrays them according to this criterion. Finally, *attribution* teases out the subset of causal conditions that can be attributed to entities with moral responsibilities or other normative connections to the outcome.

In what follows, I first recall the importance of small-N methodological approaches in policy research and situate the question of causal assessment in the context of recent developments in qualitative research methods. Next I examine the problem of indiscriminate pluralism, defining this phenomenon with greater precision than is found in the existing literature. To tackle the problem of indiscriminate pluralism, I describe a new approach to causal analysis that emphasizes causal importance. Causal importance denotes the position of an antecedent on a scale measuring its impact on a specific attribute of the outcome, assessing its contribution to the intervening causal process, or measuring additional characteristics of the antecedent itself. I discuss covariance, leverage, and attribution criteria and explore their potential value for measuring causal importance in small-N policy studies.

What's New—and Still Missing—in Small-N Research Methods

Why do small-N research strategies play such a prominent role in policy studies—a field in which researchers are generally familiar with the benefits and basic techniques of statistical analysis? Part of the answer is that policy research often deliberately focuses on small numbers of cases that show a unique departure from the norm—whether these are exemplary accomplishments or cautionary tales—and which therefore contain important lessons for the larger universe of policy practice. There is high demand for insights into the rare and inspiring accomplishments of outliers, such as successful school programs in East Harlem (Osborne & Gaebler, 1992), the community health programs of the state of Ceará, Brazil (Tendler, 1997), or Indonesia's pollution reduction efforts (BAPEDAL and World Bank, 1995). Small-N studies are also important for evaluating pilot projects that are deliberately implemented on a limited scale to allow innovators to experiment with new ideas while minimizing the cost of potential failure.[2] Policymakers and others working in the public interest want to learn about the art of the possible, and the risk of the unthinkable, not just the trend line of the probable. To learn from these experiences requires that researchers evaluate cause-and-effect relationships based on a small number of cases—be it South Africa's Truth and Reconciliation Commission or the Chernobyl nuclear accident.

Another great attraction of small-N approaches, both in theoretical and applied settings, is their ability to trace causal mechanisms.[3] The design of intelligent policy interventions requires analyses that move beyond mere patterns of correlation to include reasonably precise characterizations of the mechanisms through which posited causal variables exert their effects. Similarly, credible theories of political behavior and policy processes must not only demonstrate correlations but must establish a logic of association (George & Bennett, 2005, pp. 135–47). Yet it is widely recognized that statistical analysis, for all of its analytic power, is of limited value in tracing causal processes (Brady, 2004, pp. 267–71; Dessler, 1991; Little, 1998; Pearl, 2000, pp. 331–58; Tilly, 2001; Waltz, 1979, pp. 3–4).

The study of causal processes requires the sort of intensive, in-depth analysis that is normally only possible to undertake on a small number of cases. These are typically studied and reported as historical narratives, which when applied to causal analysis is referred to as process tracing. Process tracing may be profitably combined with large-N analyses, bringing the benefits of multiple methods to bear on questions of policy causation (Caracelli & Greene, 1993; Lin, 1998; Tashakkori & Teddlie, 2003). The important point is that even scholars who are quite comfortable with quantitative approaches often find that small-N research methods are indispensable for producing credible causal explanations (see, e.g., Baumgartner & Jones, 1993; Ostrom, 1990; Putnam, 1993).

Given the importance of small-N methods for policy studies, it is fortuitous that in recent years there has been an outpouring of literature on qualitative research methods on topics ranging from concept formation (Collier & Mahoney, 1993) to research design (Brady & Collier, 2004; King, Keohane, & Verba, 1994), measurement (Adcock & Collier, 2001; Lustick, 1996), necessary and sufficient conditions (Goertz

& Starr, 2003), and historical analysis (George & Bennett, 2005; Mahoney & Rueschemeyer, 2003). As Bennett (2003, p. 1) observes, "qualitative methods are undergoing a renaissance unlike anything seen for the last 25 years." Although most of this literature is associated with historically intensive approaches, it has developed in parallel with renewed efforts to integrate causal complexity into quantitative models (Bates, Greif, Levi, Rosenthal, & Weingast, 1998; Braumoeller, 2003; Pearl, 2000; Ragin, 1987, 2000).

What does this literature offer in the way of tools for assessing the relative contributions of component causes in small-N settings? Given that many of these studies are concerned with causation and complexity (see, e.g., Abbott, 2001; Gerring, 2005; Ragin, 1987), it is surprising how little attention has been devoted to this issue. King et al. (1994), whose insights and more controversial claims inspired much of the renaissance in qualitative methods, are typical in this regard.[4] These authors provide a wealth of useful advice on how to infer the causal effects of a given antecedent, but are silent on the question of how to rank the relative significance of several contributing causes, other than an oblique reference to the issue in a footnote concerning the futility of total historical explanation (King et al., p. 218).

Necessity and sufficiency feature prominently in the recent literature on causation (Goertz & Starr, 2003; Mahoney, 2000; Pearl, 2000, pp. 283–327) and are used in ways that imply, but do not systematically assess relative causal importance. Although they are useful devices for constructing causal explanations, necessity and sufficiency are not in and of themselves meaningful criteria for ranking causal importance. This point is missed by those who assert that an antecedent's causal significance is a function of its association with one or both of these adjectives (Hart & Honoré, 1959/1985, p. 20; Mackie, 1965; Meehl, 1977). To the contrary, to say that an antecedent is necessary or sufficient for an outcome carries no indication of its causal importance. An antecedent may be necessary but analytically unimportant, as in the role of the Big Bang as a necessary condition for the 9/11 attacks. It may likewise be sufficient but of minor importance, as in the straw that breaks the camel's back. Because the Big Bang and innumerable other conditions in the course of human evolution and social development were necessary for all contemporary political phenomena, there is no factor that is sufficient to produce any other in the abstract. Sufficiency only makes sense as a concept *given* certain background conditions. Whether or not a particular factor is sufficient to produce an outcome depends entirely on where one places the filter distinguishing background conditions from the factor of interest. Some other analytic criterion is needed to determine where to place the filter.[5]

George and Bennett (2005) make an important contribution to the study of complex causal processes in small-N research through their recent work on typological theory. Typological theory identifies categories of recurrent causal mechanisms and produces "contingent generalizations on how and under what conditions they behave in specified conjunctions or configurations to produce effects on specified dependent variables" (George & Bennett, p. 235). Closely related to this work, Ragin (1987, 2000) responds to the challenge of holism by embracing it, using Boolean algebra and fuzzy set theory to assess the consequences of different combinations of variables. But these approaches offer little guidance on how one might rank the relative contributions of several causal processes contributing to a given outcome. Ragin's work comes closest (see especially Ragin, 2006), but his procedures require relatively large numbers of cases and, in common with Goertz's (2006) application of fuzzy set analysis to necessary and sufficient conditions, he relies on a ranking criterion essentially identical to that of conventional statistics, measuring the frequency of correspondence between antecedents and outcomes.

The specter of indiscriminate pluralism has haunted some of the most thoughtful analysts in the field. Pearl (2000, p. 328) encounters the problem in his effort to operationalize causation in terms of probability theory and structural equations. Struggling with the question of why one would privilege a lit match over ambient oxygen in explaining the cause of a fire, Pearl raises (but does not address) the issue in the final passage of his magisterial work:

> The answer, I believe, lies in the pragmatics of the explanation sought . . . It appears that pragmatic issues surrounding our quest for explanation are the key to deciding which facet of causation should be used . . .

The failure of the recent literature to make significant headway on this problem comes as a surprise because pragmatic solutions to indiscriminate pluralism figured prominently in a key early precedent, Hart and Honoré's (1959/1985) classic study *Causation in the Law*, identified by Abbott (2001, p. 117) as "the greatest modern review of the practical problem of causality." Although best known for their larger effort to bridge philosophy and law, the problem of indiscriminate pluralism motivated many of Hart and Honoré's key insights. When they first stumbled across it, they identified this as "a problem scarcely mentioned before in the history of philosophy: are there any principles governing the selection we apparently make of one of a complex set of conditions as the cause?" (Hart & Honoré, 1959/1985, p. 17).

It is no coincidence that Hart and Honoré anticipated the argument presented here because they were concerned with socially relevant causal attribution in small-N settings, like a given act of murder, where one cannot resort to the Humean notion of regularity of association as a default criterion for assessing causal importance. They recognized that the way out of this problem is to construct pragmatic analytic goals—in their case, systematizing the semantics and precedents of legal jurisprudence—to ferret out and establish a hierarchy of importance among component causes.[6] Ultimately, Hart and Honoré's analysis remained confined to the courtroom and they did not push the concept of relative causal importance beyond the criterion of attribution to legally responsible agents. I will argue that progress in policy-relevant causal assessment requires that we revisit their core question and extend it to the broader concerns of policy research.

Confronting Complexity

To better understand the class of problems that this analysis is concerned with, let us consider some additional examples.

> A U.S. military transport vehicle without adequate armor inadvertently veers from a secured route and is destroyed by a roadside bomb in Iraq. The parents of a fallen soldier cry "How did this happen? Who is responsible?" How would an honest broker distinguish among the U.S. decision to invade Iraq, the activities of Iraqi insurgents, the inadequate protective equipment, and the navigation skills of the driver?
>
> A mayor reads a news article that describes a homeless man who died from hypothermia while sleeping under a bridge during an unseasonable cold spell. The man, who was found in a thin coat and was intoxicated, had a history of mental health problems and had wandered in and out of the prison and shelter systems in the absence of long-term psychiatric care. The mayor hands the news article to an aide and says, "Tell me how to make sure this doesn't happen again."
>
> Absent pressure by Bolivia's environmental groups, timber industry opposition would likely have doomed the country's exemplary forestry law.[7] These groups grew in the 1980s with help from international NGOs that flocked to Bolivia following its transition to democracy. As luck would have it, in 1993 the father of an influential environmentalist was elected President. The law passed with active support from the President, a multipartisan coalition in the congress, and a nationwide grassroots mobilization. Which of these factors were most responsible for the outcome?

Policy researchers can readily enumerate similar examples in their substantive areas of expertise. What do these scenarios have in common, and how might we make meaningful causal distinctions in these settings? At first blush these seem to share the characteristic of "complex" combinations of causes. Complexity has many dimensions, however, ranging from multiplier effects to contingency, emergent properties, nonlinear dynamics, interaction effects, and the sheer difficulty of predicting the direction of creative human endeavors (see Jervis, 1997; Tetlock & Belkin, 1996). Here I focus on scenarios in which multiple variables are each probably necessary to produce an outcome, whereas none is individually sufficient. I describe this condition as indiscriminate pluralism, lending a more precise meaning to the term than that offered by Fischer (1970, p. 175), who devised it to describe "causal explanations where the number of causal components is not defined, or their relative weight is not determined, or commonly both."[8]

These scenarios share a condition in which the antecedents' contributions to the outcome cannot be readily described in incremental, additive terms. Indiscriminate pluralism typically assumes one or a combination of three forms, summarized in Figure 1.[9] *Nested causation* describes situations in which one of the constituent causes is a necessary but insufficient precondition for another necessary cause. Often described in terms of causal chains (Goertz & Levy, in press, pp. 26–30), a representative case is the role that international NGOs played in fostering the growth of environmental groups in Bolivia, which in turn were essential for bringing about legal reforms. Nested causation does not refer to instances in which an intervening variable is wholly predicted and accounted for by the preceding cause. Whereas

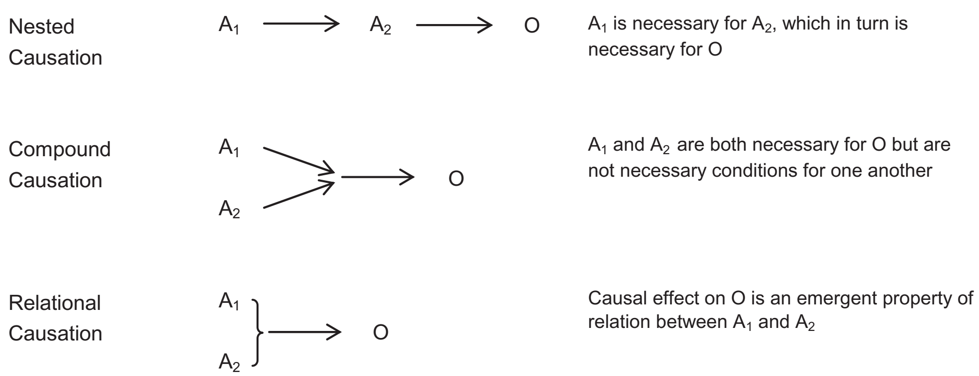

Figure 1. Types of Indiscriminate Pluralism.
A, antecedent; O, outcome.

nested causation describes asymmetrical causal relations among antecedents, a related phenomenon is what Oran Young (2002) refers to as causal clusters, in which necessary antecedents exert a mutual influence. In building political momentum for Bolivia's forestry law, for example, it is highly plausible that presidential support and the grassroots mobilization were mutually reinforcing.

Compound causation describes situations in which, in common with nested causation, multiple antecedents are each necessary but insufficient for producing a given outcome. The distinguishing feature of this category is that changing one necessary component does not change the likelihood or characteristics of the other necessary component—it only changes the *effect* of the other component. In the 9/11 case, the absence of federal air marshals had no direct bearing on the FBI's aversion to long-term investigations, but it did amplify the consequences of the FBI's orientation. Likewise, the U.S. decision to invade Iraq, the thickness of vehicle armor, and navigation skills have no effect on one another, but in tandem they amplify one another's effects on the outcome.[10]

Finally, *relational causation* describes scenarios in which the outcome is a consequence of the relative positions of contributing factors. Which was more important, the strength of the vehicle armor or the intensity of the blast? The thickness of the homeless man's coat or the unseasonably cold weather? The strength of the timber lobby or the tenacity of environmental groups? Emirbayer (1997, p. 287) contrasts these situations with the notion of interactions among variables. In relational settings, "the very terms or units involved in a transaction derive their meaning, significance, and identity from the (changing) functional roles they play within that transaction." Relational causation figures prominently in research on power dynamics and systemic theories of international relations. In expounding structural–realist theory, for example, Waltz emphasizes that the power of a given state is a function of its capabilities relative to those of other states—"how the units stand in relation to one another" (Waltz, 1979, p. 97).[11]

Indiscriminate pluralism does not portend in every case involving multiple contributing causes in small-N settings. When contributing causes are additive, for example, one can in principle readily identify their relative contributions (see Hart & Honoré, 1959/1985, pp. 225–35). If a company illegally dumps 10 tons of scrap metal in a riverbed, and a second company dumps 90 tons, their respective contributions to the outcome are not difficult to discern. However, nonadditive situations—two companies that dump reactive chemicals in the riverbed, each of which is necessary but insufficient for an explosion—are common enough that a practical solution is required. Let us consider some candidates.

Assessing Causal Importance

How can we assess the relative significance of multiple necessary causes? The intuition behind the approach described here is straightforward: causal measurement requires a metric—an axis along which one can rank the relative importance of different causal factors. Social scientists are intimately familiar with this practice for *descriptive* purposes. We might measure a bomb blast by its intensity, intoxication by blood-alcohol level, and deforestation by hectares lost per year. Likewise, concepts like political liberalization, social capital, and inter-agency cooperation are all meaningful to the extent that we can discern, in categorical or quantitative terms, when we have more or less of them. To address indiscriminate pluralism we merely need to extend this intuition to *causal* measurement. When making causal comparisons among several contributing factors, we need an axis that speaks to some trait that they have in common, but not in common measure, and can therefore be used to rank them.

The discerning reader may greet with skepticism the assertion that we need to operationalize measures of relative causal importance. How is it possible that the social sciences have not already addressed this issue, given the fundamental importance of rank ordering of causes in both theoretical and applied research? The answer is that causal importance has been construed narrowly as covariance—the degree to which a change in the antecedent is associated with a change in the outcome.[12] Covariance is such a powerful method for handling the problem of indiscriminate pluralism—and so conveniently corresponds to the ontology and practice of regression analysis (see Hall, 2003)—that its habitual application has invited premature closure on the meaning and operationalization of causal importance. Yet, as I discuss below, covariance approaches are often infeasible or inappropriate in small-N settings. This presents an opportunity to revisit the larger analytic question to which covariance is one answer and to explore whether other answers might be found.

In doing so I will rely heavily on the concept of causal importance—a notion that is often implied in discussions of causation yet has not been a focus of explicit attention in the literature. Causal importance refers to the position of an antecedent on a scale derived from a pragmatic analytic goal. Logically and semantically we can only say that an antecedent is a more or less important cause of an outcome "with respect to" such a goal. Assessment of causal importance has three variants. With *outcome-based* causal assessment, the investigator constructs a metric with respect to the outcome—emphasizing its timing, extent, frequency, or other characteristics—and then analyzes the degree to which one or another antecedent is responsible for that dimension of the outcome. In *process-based* causal assessment, one applies the same tool to the intervening processes linking antecedents and outcomes, identifying key components of the causal mechanisms at play, and ranking antecedents according to their contributions to those components. In *antecedent-based* causal assessment, the researcher constructs a metric for ranking other characteristics of the causally relevant antecedents themselves, such as their susceptibility to manipulation. Thus if two antecedents make essentially similar contributions to the outcome and to the intervening process, we are not out of options for making distinctions with respect to their causal importance. In what follows I begin by discussing the advantages and drawbacks of covariance analysis, which is the most popular approach for outcome-based causal assessment. I then describe a number of alternatives, including disaggregation techniques (applicable to both outcome- and process-based assessment) and antecedent-based assessment strategies that rank antecedents according to their susceptibility to leverage and their standing with respect to criteria for moral, legal, and other forms of normative attribution.

Covariance

Covariance analysis singles out those antecedent conditions for which a change in value is associated with a change in the outcome. Two elements in this description are especially noteworthy because together they account for the approach's key strength and its principal limitation. First, covariance approaches disregard any antecedents that do not vary across cases. Second, they attempt to explain only the marginal difference in outcomes across the cases, rather than offer a more global assessment of the origins of those outcomes. In other words, covariance also disregards any characteristics of the outcome that do not vary across the cases under investigation.

Through these two characteristics, covariance approaches resolve in one swift blow the problem of indiscriminate pluralism. If an investigator were to compare a number of Moroccan villages to better understand Ait Iktel's success, covariance analysis would focus on its marginal difference from the other villages, disregarding (i.e., controlling for) the effect of Moroccan liberalization and that of any other condition (the Big Bang, norms of village reciprocity) that is constant across cases. In large-N settings, covariance permits discrimination among component causes by comparing the average effect that a change in one antecedent variable has on an outcome relative to the average effect of changing another antecedent. Presented in convenient formats like ANOVA tables and standardized partial correlation coefficients, large-N covariance approaches neatly solve the problem.

Covariance analysis is so powerful that it has become identified in the social science imagination as synonymous with causal importance, subsuming the latter and obviating the need for its elaboration as a distinct and perhaps farther reaching concept.[13] To appreciate the shortcomings of the covariance approach, however, let us return to the role of inadequate information sharing among federal agencies as a necessary condition for the 9/11 attacks, insofar as information exchanges might have been sufficient to derail al Qaeda's plans (Kean & Hamilton, 2004, pp. 592–96). In contrast to within-case methods such as process tracing, a covariance approach requires additional cases for comparison. For U.S. policymakers, the variation of interest is the difference between domestic conditions of peace and security and the 9/11 attacks, so let us compare the United States in August 2001 to the United States in August 1991.

This produces the following set of correlations: no attacks in 1991 in the presence of poor information sharing among agencies, and attacks in 2001 in the presence of poor information sharing. Because the antecedent condition of inadequate information sharing does not vary across these cases, in covariance terms it is irrelevant. This impression is cemented by the unfortunate terminology of covariance analysis, which designates as "trivial" any conditions that do not vary across cases (Braumoeller & Goertz, 2003, pp. 219–23; Goertz, 2006). Using covariance as the criterion for ordering causes, we would focus on the conditions that did vary across cases—such as the rise of al Qaeda—as the only nontrivial causes of the outcome on 9/11, thereby missing most of the other causes listed in Table 1. If we focus our attention on the domestic conditions that enabled these attacks (the aspect of greatest interest to the 9/11 Commission), a covariance-oriented approach offers very little, as there were no substantial changes in domestic conditions correlated with the change in outcome. From a policy perspective, however, one is often concerned with existing conditions that do not vary across cases in practice but could do so in the future as a result of intervention. Note that the constant conditions in the 9/11 example are not immutable social structures; the very factors that do not vary in practice could vary in principal.

From a practical standpoint, covariance approaches are often ill suited to causal assessment in small-N settings. Covariance is grounded in Hume's notion of regularity of association. In small-N settings, covariance across cases does not occur with the sort of regularity that Hume had in mind (and that statistics and experimental control achieve in practice) for drawing valid causal inferences. If a second Moroccan village were discovered to have social accomplishments equal to those of Ait Iktel absent its social capital, or equal social capital without similar accomplishments, this would in no way disprove the causal importance of social capital in promoting rural development in Ait Iktel or elsewhere. Only with the most extreme assumption of deterministic causation could this be the case (see George & Bennett, 2005, p. 163; Lieberson, 1991). This type of controlled comparison of cases, originating in J. S. Mill's (1843/1967) work, is nominally inspired by experimental methods. But experimental control as practiced in the natural sciences is a large-N method because experimental scientists recognize that most causal relations of interest are probabilistic and therefore require repeated tests. Thus even when medical researchers have at their disposal thoroughly inbred, genetically identical mice—a similarity of background conditions unimaginable in comparative social inquiry—they still use large numbers of such mice before inferring the treatment effect of an experimental drug.

As an alternative to covariance analysis, the influence of a factor like inter-agency cooperation in the 9/11 case can be revealed using within-case methods such as process tracing. Process tracing is defined by George and McKeown (1985, pp. 35–36) as studying the "process by which various initial conditions are translated into outcomes." In contrast to covariance approaches, process tracing evaluates "a stream of behavior through time . . . Any explanation of the processes at work in the case thus not only must explain the final outcome, but also must account for this stream of behavior." Process tracing and covariance analysis share some elements in common. They rely on the same definition of causation, namely the effect of an antecedent on an outcome (see Brady, 2003; Gerring, 2005).[14] And the need to understand the causal mechanisms through which antecedents produce these effects is a matter of far greater consensus than is suggested by debates on the relative merits of studying effects versus mechanisms. The difference lies in the method for analyzing these phenomena. With process tracing, rather than drawing causal inferences from patterns of association between antecedents and outcomes, the investigator assesses the *logic* of the association. One examines not merely the presence or absence of an antecedent like poor inter-agency cooperation, but the specific ways in which this made the 9/11 attacks possible, such as the FBI's failure to share information about suspected terrorists taking flight lessons. The investigator breaks down complex chains of events into smaller pieces, and distant relations between antecedents and outcomes into more proximate cause-and-effect couplings. Causal processes are selectively decomposed further and further until the plausibility of the component cause-and-effect relationships is so high that further explanation is unwarranted (Roberts, 1996).

This alternative, historically intensive approach to causal assessment comes at a cost, however, because the great advantage of process tracing—its ability to produce rich causal narratives—gives rise to a significant weakness, as process tracing carries none of the built-in remedies for indiscriminate pluralism that accompany covariance approaches like regression analysis. Let us then consider some additional options for assessing causal importance in small-N research.

Disaggregation

The widespread use of the terms *necessity* and *sufficiency* contributes significantly to the problem of indiscriminate pluralism by transforming what are typically multifaceted, cumulative outcomes into simpler, dichotomous (presence/absence) categories of outcomes that are easily grasped yet resist assessments of relative causal importance. Categorical descriptions are pervasive in everyday conversation and in scholarly discourse, suggesting that they play an important role in our ability to render a complex world intelligible. Yet expressions that imply necessary conditions—such as "The President's support was essential for the law's passage," "The project's success relied on a centuries-old tradition of village reciprocity," and

"What made the country vulnerable?"—always imply dichotomous dependent variables (passage/no passage, success/failure, vulnerable/not vulnerable).[15] In principal, necessity and sufficiency are compatible with cumulative measurements (Necessary for one? Necessary for two? Were two of these necessary for three of those?) and can even be operationalized in probabilistic terms when enough cases are available (Ragin, 2000). But in practice, the semantics of necessary and sufficient conditions create a habit of describing the world in terms of categories and thresholds, which, by virtue of their simple and dichotomous nature, render it more difficult to identify and compare the incremental contributions of component causes.

The use of categorical dependent variables is less problematic in large-N assessments, wherein one can measure the extent to which changes in independent variables make a sizeable number of discrete outcomes more likely; the resulting quantitative distinctions offer a ready metric for assessing relative causal contributions, whether through regression models for categorical and limited dependent variables (Long, 1997) or through fuzzy set analysis (Goertz, 2006; Ragin, 2006).

For small-N research, disaggregation of dichotomous variables into their component parts is a viable alternative. King et al. (1994, p. 217) prescribe disaggregation of dependent variables as a method for "making many observations from few," thereby rendering small-N studies more amenable to covariance analysis. They recommend, for example, studying geographical subunits of a national phenomenon or particular agencies or decisions that serve as a test for hypotheses applied to the state as a whole. Whereas King and colleagues prescribe disaggregation for increasing the validity of causal inferences, this same prescription can cure a different malady, that of indiscriminate pluralism. Disaggregation offers a way out of the paradox of the relative importance of dried twigs, ambient oxygen, and a tossed match in causing a fire. The problem lies in the description of the outcome in dichotomous terms (fire/no fire). If we consider instead a specific attribute such as the intensity of the fire—was it a minor event or a major conflagration?—the quantity of dried twigs might prove more important with respect to that dimension of the outcome. Likewise the categorical event of "the 9/11 attacks" can be disaggregated into the timing, location, degree of destructiveness, vulnerability of urban centers, choice of weapons, the first plane attack versus subsequent attacks, and other dimensions along which one can make a number of meaningful distinctions regarding the relative contributions of al Qaeda and the factors shown in Table 1. In the case of Ait Iktel, the village's "success" can be disaggregated into facets such as improved income generating activities for women, for which some antecedents (such as a specific donor project) would likely prove more important than others. If desired, these more tractable pieces of causal analysis can then be re-aggregated to reach summary conclusions regarding the relative importance of factors like external financial support and local social entrepreneurs.

Disaggregating a phenomenon into smaller pieces is the very essence of process tracing, which dissects not the outcome itself but the mechanisms through which antecedents influence that outcome. This affords the opportunity for a second approach to causal assessment—process-based assessment—in which the investigator ranks the relative importance of antecedents with respect to their impact on the intervening processes. In the 9/11 case, the investigator could identify a dozen or more stages in the planning of the attacks and might discover that information exchanges among government agencies could have thwarted the plot during several of these stages. In contrast, the absence of federal air marshals only enabled the final stage of the attack and was therefore less causally important with respect to this measure. Alternatively, one can identify critical junctures in a causal pathway—key moments in the evolution of a project or policy process that determine its future viability and orientation—and trace the extent to which one or another antecedent was responsible for these critical junctures. For example, the first foreign donors to support Ait Iktel (such as the French government and Moroccans living abroad) likely enhanced the village's ability to attract funds from subsequent donors. These first movers were therefore arguably a more important cause of success than those donors who jumped on the bandwagon after the village's first accomplishments attracted attention.

Leverage

In contrast to outcome- and process-based assessment, antecedent-based assessment measures additional characteristics of the antecedents themselves, filtering out those antecedents irrelevant to the attribute in question and arraying those that remain. To illustrate, Table 2 ranks the relative importance of several documented causes of Ait Iktel's success according to alternative metrics of leverage (susceptibility to manipulation) and attribution (in this case, who deserves credit) and compares these rankings to that produced by a covariance approach. As in descriptive comparison, the rank of causal importance assigned to a given phenomenon will differ according to the measurement employed. Just as the relative descriptive ordering of a bowling ball and a feather will be reversed depending on the characteristic measured—weight or suitability for pillow stuffing—so too do the causal ranking of political liberalization and local entrepreneurship differ according to the causal

Table 2. Ranking Causes of Ait Iktel's Success

Causal Importance	Measurement Attributes		
	Leverage[a]	Attribution[b]	Covariance[c]
Very important	External finance	External finance Local entrepreneurs Political liberalization[d]	External finance Local entrepreneurs
Somewhat important	Growth of NGOs Local entrepreneurs		
Unimportant	Social capital Political liberalization	Social capital Growth of NGOs	Social capital Political liberalization Growth of NGOs

[a]How might international development agencies promote similar outcomes elsewhere?
[b]Which normatively relevant actors deserve credit for the outcome?
[c]What accounts for the difference between Ait Iktel and neighboring villages?
[d]King Hassan II's role in particular is pertinent to attribution.

attribute measured. This in no way implies that physical reality is purely in the eye of the beholder. It is a question of making that reality intelligible through the isolation and measurement of salient traits.

Among potential measurement criteria, leverage stands out as especially relevant to the policy sciences. Sabatier (1991, pp. 144–45) notes that the origin of public policy as an academic field can be traced in part to "pressures to apply the [political science] profession's accumulated knowledge to the pressing social problems of racial discrimination, poverty, the arms race, and environmental pollution . . ." Moreover, "Most policy scholars have an activist bent, i.e. at some point they wish to influence policy in the area(s) in which they are specialists" (see also Lindblom & Cohen, 1979; Steinberg, 2005). Using leverage as a metric of causal importance, the investigator brings to the foreground those antecedents subject to manipulation and ranks them according to their perceived susceptibility to intervention, which may include assessment of cost effectiveness and will typically differ from one agent of change to the next according to their capabilities and preferred strategies.

Leverage requires a move from explanation to prescription—from why, to how, to how to. Especially in small-N research designs, this often involves a prescription for cases other than those studied, requiring assessment of actual causal importance in the cases at hand and potential causal importance in other times, places, and conditions (Rose, 1993). Given the complexity of the causal narratives resulting from process tracing, the investigator must distill some tractable subset of antecedents that are generalizable (achievable under a wider range of conditions) yet sufficiently precise in their effects to produce the desired outcome. What measurement criteria might we apply to privilege certain antecedents over others for this purpose? One useful approach, especially in cases characterized by nested causation, is to envision a sliding lens that can be placed anywhere along the stream of events leading to an outcome, with "downstream" representing its proximate causes and "upstream" the more distant preconditions for these causes. The investigator rests the lens on that part of the causal stream where fulfillment of the conditions present at that point would increase the probability of the desired result to an acceptable level. The analysis then requires looking upstream to discern the variables (structural or otherwise) needed to arrive at that point in the causal stream, focusing not just on those necessary in the case at hand, but on a broader range of possibilities.

Let us apply this approach to the Ait Iktel case. Here we have a setting characterized by compound causation—with social capital, external finance, and local social entrepreneurs all necessary for the outcome—and by nested causation, as political liberalization was a prerequisite for the rise of NGOs, which in turn helped these entrepreneurs to mobilize people and resources. To reiterate, the position of the sliding lens is balanced by two competing pressures: acceptable likelihood of results and generalizability to other potential cases. The first pressure, in which acceptability is a function of the aspirations of interveners relative to the difficulty and uncertainty of the outcome, moves the lens downstream in the direction of greater proximity to the outcome. If the lens were placed too far upstream—focusing, for example, on political liberalization and steps to promote it—the downstream result of effective village-level social services would be so uncertain that this intervention would not produce an acceptable likelihood of results by the standards of most development advocates. The second, competing pressure—the desire for generalizability to other potential cases—tends to move the lens upstream. This is a familiar practice in theory building, as proper nouns are removed and concepts are stretched to greater levels of abstraction in order to apply to a broader range of settings (Collier & Mahoney, 1993; Sartori, 1970). Here one might choose to downplay the necessity of fundraising skills on the part of village leaders (an actual proximate cause in Ait Iktel) if it is judged that there are other feasible ways to make external resources available to local grassroots organizations. In light of these competing pressures, the sliding lens might then settle on "local social capital mobilized by dynamic village leaders who have the right to organize and have access to external financial resources."

Attribution

Attribution is another metric for causal assessment which, like a magnet swept across a cluttered workbench, selectively draws out a collection of related items from an otherwise jumbled mixture. Attribution-oriented causal assessment can take a number of forms. One may wish to assign praise or blame for an outcome. Alternatively, the focus might be on program evaluation, highlighting the subset of causal pathways shaped by an intervention of interest and comparing the actual versus intended effects. Impact studies emphasize yet another dimension of attribution, as the investigator traces not only the intended causal pathway but also the broader social consequences of new developments, such as the environmental impacts of a free trade agreement (CEC, 1999) or the economic cost of a new environmental regulation (Jaffe, Peterson, Portney, & Stavins, 1995).

Attribution lies at the core of Hart and Honoré's (1985, p. 24) analysis. They argue that "the source of the lawyer's main perplexities" in causal explanation

> . . . are generated less by ignorance of fact than by the vagueness or indeterminacy of the very concept of causal connection which we are endeavoring to apply in a particular case. Typically what precipitates these difficulties is that, among the conditions required to account for the harm which has occurred, there is found in addition to the defendant's action a factor (usually a human action or some striking natural phenomenon) which itself has some of the characteristics by which common sense distinguishes causes . . .

Although policy researchers share some of the attributive interests of the legal profession, there are essential differences as well. Lawyers put drug dealers and terrorists on trial; they do not put on trial the social conditions (such as the efficacy of existing laws!) that render phenomena like drug abuse and terrorism more or less likely. In other words, there are entire categories of policy-relevant attribution that are not captured by discussions of legal liability. Responsibility for trends in homelessness, rural electrification, military casualties, forest conservation, and other outcomes cannot be sorted using legal criteria alone.

One place to look for policy-relevant attribution criteria is in the normative obligations embedded in institutions. Whether they take the form of laws, regulations, constitutions, organizations, or rules in use, institutions carry norms and associated social expectations that offer a basis for ranking causal importance. Institutional norms are encoded in rules that assign different normative obligations to different actors according to their roles. The distance between actual performance and the normative expectations stemming from institutional roles can serve as a metric for ranking causal importance. Actors who failed to meet their responsibilities or, alternatively, went "beyond the call of duty" in promoting an outcome may be judged more important causes of that outcome than others who were just doing their jobs, because the former constitute extraordinary circumstances. (On the abnormality criterion, see Goertz, 2006, pp. 12–14; Hart & Honoré, 1959/1985, pp. 162–85, 340–62; Roberts, 1996, pp. 96–99.) For example, in assessing causes of American vulnerability to the 9/11 attacks, legal mandates and associated social expectations designate some actors, such as the CIA and FBI, as more normatively connected than others to counterterrorism efforts, and hence their actions emerge as more causally important than those of city officials who may have been inadequately prepared for such an attack.

The distance between performance and normative aspirations is central to program evaluation, which might appear to circumvent the problem of indiscriminate pluralism by highlighting only those effects directly related to the intervention. As Chen (1990) has argued, however, too often program evaluation suffers from the opposite problem, tracing antecedent impacts without paying adequate attention to the broader causal processes driving the social problem in question. This can be seen in evaluations of the impact of international environmental institutions on national outcomes. Steinberg (2001, pp. 203–6) argues that these studies typically fail to consider models of policy reform and social change in the target countries, leading to overestimation of the effect of international advocacy campaigns and underestimation of the impact of international aid. Here the problem is not the result of over-explanation and indiscriminate pluralism, but of focusing in so tightly on a given antecedent that the broader array of causal forces is missed. Thus, while this discussion has focused on the pitfalls of holism and the challenges of causal assessment in complex settings, it is better to reveal complexity and to manage it than to not engage complexity at all.

Conclusion

Paul Rich (2002, p. 530) has argued in the pages of this journal that "We need to see more discussion in policy studies of the various tools of the trade, and history is one of them." While few political scientists would dispute the importance of analyzing historical processes, there is little agreement on how to do it well or on the larger question of its role in producing verifiable propositions about cause and effect. Process tracing has emerged as a promising tool for combining the historian's craft with the political scientist's commitment to the systematic evaluation of causal claims. Yet it would appear that the greater the informational richness revealed by process tracing, the higher the risk of producing analytically unwieldy explanations, as the resulting causal narratives reveal numerous necessary and socially relevant causes with no clear means for discerning their relative importance.

Indiscriminate pluralism poses a serious challenge for small-N research because the comparative advantage touted by proponents of process tracing—the ability to capture complex social processes—is the very trait that produces potentially confounding explanations. Small-N research encounters this problem not because the subjects of quantitative analysis are characterized by lower complexity—they most certainly are not—but because statistical tools neatly circumvent indiscriminate pluralism by reducing the research question to an analysis of covariance. Solving this problem in small-N research is not as simple as borrowing the tools of covariance analysis because in small-N settings this produces only weak causal inferences. Transforming an investigation into a large-N study may make sense in some settings, but is often infeasible, as in the 9/11 commission report, or inappropriate, when causal mechanisms or noteworthy outliers are of special interest. Historical process tracing and other forms of within-case analysis are here to stay. So too are their inherent challenges.

This article has presented a number of strategies for addressing these challenges, founded on the concept of causal importance—the position of antecedents on a scale describing some characteristic of the outcome, of the process, or of the antecedents themselves. This approach can be applied at various stages of the research process. With respect to the formation of research questions, I argue that the investigator seeking to identify the most important causes of an outcome is posing a poorly crafted question. One must clarify, important with respect to what attributes of the causal relationship? A common response to the graduate student who poses a seemingly open-ended causal question is to advise that he or she focus on explaining patterns of covariance, perhaps through a controlled comparison of cases or regression analysis. This may be entirely appropriate in some circumstances, but a premature leap from question formation to design strategy risks obscuring and even distorting the underlying purpose of the study. The preceding analysis of the 9/11 and Ait Iktel cases demonstrates that, from the standpoint of policy relevance, it is often just as important to examine the impact of antecedents that do not vary across cases. By focusing greater attention on the process of selecting relevant causal attributes, an emphasis on causal importance encourages more careful reflection on the normative underpinnings of research questions and helps to ensure that the purpose of the study drives the research design, rather than the reverse.

In the course of field research, the practice of process tracing typically reveals many causal factors beyond what the investigator can reasonably anticipate during the formation of research questions. I have described a number of approaches for rendering complex causal narratives more manageable, most notably through the identification of attributes that can be used to rank the relative importance of contributing causes, as well as specific techniques such as disaggregation and the sliding lens approach. Measurements of causal importance cannot necessarily untangle every strand of indeterminacy in a given scenario, but go a long way toward rendering these problems more analytically tractable. In some cases the investigator may deem that the importance of two or more factors is essentially equivalent with respect to a given category of measurement, as in the ranking of local entrepreneurs and NGO growth in the leverage column in Table 2. This is distinct from and highly preferable to situations of indeterminacy, in which causally relevant antecedents have equal standing simply because we lack a method for discerning among them.

In the process of identifying relevant causal attributes, one can profitably combine several of the measurement criteria described here. Judgment of an antecedent's importance with respect to attribution or leverage will often be shaped in part by its perceived impact on the outcome and on the intervening causal process. There are, moreover, points of contact between leverage and attribution criteria: Laws are often designed to not only punish offenders but also to leverage changes in others' behavior (Calabresi, 1975, cited in Hart & Honoré, 1959/1985, p. lxxii). Likewise, judgment regarding an agency's liability for an undesirable outcome may hinge on perceptions of its potential leverage with respect to that outcome. There are no doubt many other policy-relevant attributes beyond those presented here that could be applied within the general framework of assessing causal importance.

The concept of causal importance has much to recommend the field of policy studies, mirroring the field's unique pairing of explicitly normative aspirations with a commitment to objective (i.e., inter-subjectively verifiable) analysis. When clarified and made explicit, these subjectively derived metrics of causal importance are not only compatible with scientific inquiry into cause-and-effect relationships, they are a prerequisite for it. It is only when causal importance is mistakenly assumed to have substantive meaning in the absence of measurement scales that we run into problems, either comparing incomparables or using covariance as the default measure of importance. Researchers who focus on policy-relevant aspects of a causal relationship are not engaging in a parochial departure from the broader goal of social science explanation. To search for the "real" causes of an outcome absent a metric for causal importance is as futile as searching, in descriptive settings, for the true essence of an object. Characterizations require characteristics, and in causal analysis the relevant traits reside in the outcome, in the antecedents, in the relation between the two, and in their connection to pragmatic analytic goals.

Paul F. Steinberg is assistant professor in Political Science and Environmental Policy at Harvey Mudd College.

Notes

For insightful comments on earlier drafts of this article, I wish to thank Eugene Bardach, George Busenberg, Gary Goertz, James Mahoney, Ronald Mitchell, Craig Thomas, and three anonymous reviewers from the Policy Studies Journal. This research was supported in part by a grant from the Andrew W. Mellon foundation.

1. On the considerable advantages of construing causation in probabilistic terms, see Pearl (2000, pp. 1–2).
2. Statistical methods can of course be used to evaluate a pilot project implemented in even just one city if the results can be measured as individual effects (behavior change, customer satisfaction) in a large

population. In contrast, understanding how the pilot project influenced and was shaped by political sovereigns and agency leaders in that city requires intensive, small-N techniques.

3. The concept of mechanisms used here is consistent with that of Tilly (2001) and Little (1998), rather than the methodological individualism associated with Hedström and Swedberg (1998), who equate mechanisms with individual decisions. The latter approach may be appropriate in some settings, but given the emergent properties of social systems (and of complex systems generally) there is no compelling reason to insist a priori on recourse to the smallest decision-making unit.
4. For a comprehensive and constructive critique of *Designing Social Inquiry*, see Brady and Collier (2004).
5. Mahoney (2003, pp. 348–50) observes that researchers typically use necessity and sufficiency in ways that are not trivial, and I make no claim to the contrary. Rather I am arguing that while these concepts can be useful in constructing metrics for ranking causal importance, they do not constitute metrics in and of themselves.
6. Gerring's criterial approach to causal assessment shares some elements in common with Hart and Honoré but focuses on the merits of causal *explanations* and therefore includes criteria such as parsimony and innovation, whereas Hart and Honoré, and the present analysis, focus on the importance of causal forces themselves. There is some overlap with Gerring's approach when the merits of particular explanations are a function of the plausibility of their posited causes (see Gerring, 2005, pp. 175–79, 187–88).
7. The Bolivian example is from Steinberg (2001).
8. Many authors refer loosely to these situations as "overdetermined." The term is not adequate for the present purpose because the "over" in overdetermination stems from a constraint unique to quantitative methods, namely the mathematical requirements for solving simultaneous equations.
9. The categories described here share some elements in common with more general typologies of causation (see for example Jervis, 1997, pp. 34–60) but differ in focus, identifying scenarios that produce indiscriminate pluralism. For an instructive exercise in diagramming narrative arguments that rely on multiple necessary conditions, see Mahoney (in press).
10. This is distinct from the Bayesian notion of independence, in which knowledge of one antecedent has no effect on our *understanding* of the other. Arguably, knowledge about armor thickness tells us something about U.S. preparedness, which could inform predictions regarding navigation training of drivers. Hence the two antecedents are independent with respect to compound causation, but not with respect to Bayesian approaches.
11. Note that interaction variables in regression analysis do not address relational causation because they do not measure the interacting variables' relative contributions to the interaction. If we include in the model both the individual variables and the interaction term, the results measure the former's independent contribution to the outcome apart from any interaction effects, not their relative importance in the interaction.
12. I use the term "covariance" rather than its popular shorthand "variance" to avoid potential ambiguity associated with the latter. All "why" questions imply a kind of variance—why this and not that, why here and not there, why now and not then? Likewise any event or noteworthy condition implies variance from baseline conditions. Covariance is distinct from these broader uses of variance, denoting a condition of correspondence between changing antecedents and changing outcomes.
13. Pearl (2000, pp. 331–58) provides a fascinating historical account of how the development of modern probability and statistics precluded a language appropriate for causal assessment.
14. I am indebted to an anonymous reviewer for bringing this issue to my attention.
15. These are also referred to as nominal or discrete dependent variables.

References

Abbott, Andrew. 2001. *Time Matters: On Theory and Method*. Chicago: University of Chicago Press.

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard For Qualitative and Quantitative Research." *American Political Science Review* 95: 529–46.

BAPEDAL and World Bank. 1995. *What Is PROPER? Reputational Incentives for Pollution Control in Indonesia*. Washington, DC: BAPEDAL and World Bank.

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

Baumgartner, Frank, and Bryan D. Jones. 1993. *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.

Bennett, Andrew. 2003. "Letter from the Transitional President." *Qualitative Methods* 1 (1): 1.

Brady, Henry E. 2003. "Models of Causal Inference: Going Beyond the Neyman-Rubin-Holland Theory." Presented at the annual meeting of the Midwest Political Science Association, Chicago.

———. 2004. "Data-Set Observations versus Causal-Process Observations: The 2000 U.S. Presidential Election." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, ed. Henry E. Brady, and David Collier. Lanham, MD: Roman and Littlefield, 267–71.

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

Braumoeller, Bear F. 2003. "Causal Complexity and the Study of Politics." *Political Analysis* 11 (3): 209–33.

Braumoeller, Bear F., and Gary Goertz. 2003. "The Statistical Methodology of Necessary Conditions." In *Necessary Conditions: Theory, Methodology, and Applications*, ed. Gary Goertz, and Harvey Starr. Lanham, MD: Roman and Littlefield, 197–233.

Calabresi, Guido. 1975. "Concerning Cause and the Law of Torts." *University of Chicago Law Review* 43: 69–108.

Caracelli, Valerie J., and Jennifer C. Greene. 1993. "Data Analysis Strategies for Mixed-method Evaluation Designs." *Educational Evaluation and Policy Analysis* 15: 195–207.

Chen, Huey-Tsyh. 1990. *Theory-Driven Evaluations*. Newbury Park, CA: Sage Publications.

Collier, David, and James Mahoney. 1993. "Conceptual Stretching Revisited: Adapting Categories in Comparative Analysis." *American Political Science Review* 87 (4): 845–55.

Commission for Environmental Cooperation (CEC). 1999. *Assessing the Environmental Effects of the North American Free Trade Agreement: Final Analytic Framework and Background Document*. Montreal, QC: CEC.

Dessler, David. 1991. "Beyond Correlations: Toward a Causal Theory of War." *International Studies Quarterly* 35: 337–55.

Emirbayer, Mustafa. 1997. "Manifesto for a Relational Sociology." *American Journal of Sociology* 103 (2): 281–317.

Fischer, David H. 1970. *Historians' Fallacies: Toward a Logic of Historical Thought*. New York: Harper & Row.

George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.

George, Alexander L., and Timothy J. McKeown. 1985. "Case Studies and Theories of Organizational Decision Making." In *Advances in Information Processing in Organizations*. Vol. 2, ed. Robert F. Coulam, and Richard A. Smith. Greenwich, CT: JAI Press, 21–58.

Gerring, John. 2005. "Causation: A Unified Framework for the Social Sciences." *Journal of Theoretical Politics* 17 (2): 163–98.

Goertz, Gary. 2006. "Assessing the Trivialness, Relevance, and Relative Importance of Necessary or Sufficient Conditions in Social Science." *Studies in Comparative International Development* 41 (2): 88–109.

Goertz, Gary, and Harvey Starr, eds. 2003. *Necessary Conditions: Theory, Methodology, and Applications*. Lanham, MD: Roman and Littlefield Publishers.

Goertz, Gary, and Jack S. Levy. in press. "Causal Explanation, Necessary Conditions, and Case Studies." In *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*, ed. Gary Goertz, and Jack S. Levy. New York: Routledge, 9–40.

Hall, Peter A. 2003. "Aligning Ontology and Methodology in Comparative Research." In *Comparative Historical Analysis in the Social Sciences*, ed. James Mahoney, and Dietrich Rueschemeyer. New York: Cambridge University Press, 373–404.

Hart, H. L. A., and Tony Honoré. 1959/1985. *Causation in the Law*. 2nd ed. New York: Oxford University Press.

Hedström, Peter, and Richard Swedberg, eds. 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. New York: Cambridge University Press.

Jaffe, Adam B., Steven R. Peterson, Paul R. Portney, and Robert N. Stavins. 1995. "Environmental Regulation and the Competitiveness of U.S. Manufacturing: What Does the Evidence Tell Us?" *Journal of Economic Literature* 33 (1): 132–63.

Jervis, Robert. 1997. *System Effects: Complexity in Political and Social Life*. Princeton, NJ: Princeton University Press.

Kean, Thomas H., and Lee H. Hamilton. 2004. *The 9/11 Report: The National Commission on Terrorist Attacks Upon the United States*. New York: St. Martin's Press.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Lazarsfeld, Paul F., and Allen H. Barton. 1951. "Qualitative Measurement in the Social Sciences: Classification, Typologies, and Indices." In *The Policy Sciences*, ed. Daniel Lerner, and Harold D. Lasswell. Stanford, CA: Stanford University Press, 155–92.

Lieberson, Stanley. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70 (2): 307–20.

Lin, Ann Chih. 1998. "Bridging Positivist and Interpretivist Approaches to Qualitative Methods." *Policy Studies Journal* 26 (1): 162–80.

Lindblom, Charles E., and David K. Cohen. 1979. *Usable Knowledge: Social Science and Social Problem Solving*. New Haven, CT: Yale University Press.

Little, Daniel. 1998. *Microfoundations, Method, and Causation*. New Brunswick, NJ: Transaction Press.

Long, Scott J. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

Lustick, Ian S. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias." *American Political Science Review* 90 (3): 605–18.

Mackie, John L. 1965. "Causes and Conditionals." *American Philosophical Quarterly* 2: 245–65.

Mahoney, James. 2000. "Strategies of Causal Inference in Small-N Analysis." *Sociological Methods and Research* 28 (4): 387–424.

———. 2003. "Strategies of Causal Assessment in Comparative Historical Analysis." In *Comparative Historical Analysis in the Social Sciences*, ed. James Mahoney, and Dietrich Rueschemeyer. New York: Cambridge University Press, 337–72.

———. in press. "The Elaboration Model and Necessary Causes." In *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*, ed. Gary Goertz, and Jack S. Levy. New York: Routledge, 235–61.

Mahoney, James, and Dietrich Rueschemeyer, eds. 2003. *Comparative Historical Analysis in the Social Sciences*. New York: Cambridge University Press.

Meehl, Paul E. 1977. "Specific Etiology and Other Forms of Strong Influence: Some Quantitative Meanings." *Journal of Medicine and Philosophy* 2: 33–53.

Mernissi, Fatema. 1997. "Social Capital in Action: The Case of the Ait Iktel Village Association." In *Voices from Marrakech: Towards Competitive and Caring Societies in the Middle East and North Africa*, ed. Ishac Diwan, and Karen Sirker. Washington, DC: The World Bank.

Mill, John Stuart. 1843/1967. *A System of Logic: Ratiocinative and Inductive*. Toronto, ON: University of Toronto Press.

Osborne, David, and Ted Gaebler. 1992. *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector*. Reading, MA: Addison Wesley.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Pressman, Jeffrey L., and Aaron Wildavsky. 1973. *Implementation*, 2nd ed. Berkeley: University of California Press.

Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

Ragin, Charles C. 1987. *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.

———. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

———. 2006. "Set Relations in Social Research: Evaluating Their Consistency and Coverage." *Political Analysis* 14: 291–310.

Rich, Paul. 2002. "Policy Studies and History." *Policy Studies Journal* 30 (4): 528–31.

Roberts, Clayton. 1996. *The Logic of Historical Explanation*. University Park: Penn State University Press.

Rochon, Thomas R., and Daniel A. Mazmanian. 1993. "Social Movements and the Policy Process." *The Annals of the American Academy of Political and Social Sciences* 528: 75–87.

Rose, Richard. 1993. *Lesson-Drawing in Public Policy*. Chatham, NJ: Chatham House Publishers.

Sabatier, Paul A. 1991. "Toward Better Theories of the Policy Process." *PS: Political Science and Politics* 24: 147–56.

———. 1999. *Theories of the Policy Process*. Boulder, CO: Westview Press.

Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64 (4): 1033–53.

Steinberg, Paul F. 2001. *Environmental Leadership in Developing Countries: Transnational Relations and Biodiversity Policy in Costa Rica and Bolivia*. Cambridge, MA: MIT Press.

———. ed. 2005. "Is Anyone Listening? The Impact of Research on Global Environmental Practice." Special Issue, *International Environmental Agreements: Politics, Law, and Economics* 5 (4): 377–79.

Tashakkori, Abbas, and Charles Teddlie, eds. 2003. *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, CA: Sage Publications.

Tendler, Judith. 1997. *Good Government in the Tropics*. Baltimore, MD: Johns Hopkins University Press.

Tetlock, Philip E., and Aaron Belkin. 1996. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton, NJ: Princeton University Press.

Tilly, Charles. 2001. "Mechanisms in Political Processes." *Annual Review of Political Science* 4: 21–41.

Waltz, Kenneth N. 1979. *Theory of International Politics*. New York: Random House.

World Bank. 2003. *World Development Report 2003*. New York: Oxford University Press.

Young, Oran R., ed. 1999. *The Effectiveness of International Environmental Regimes: Causal Connections and Behavioral Mechanisms*. Cambridge, MA: MIT Press.

———. 2002. "Are Institutions Intervening Variables or Basic Causal Forces? Causal Clusters versus Causal Chains in International Society." In *Millennial Reflections on International Studies*, ed. Michael Brecher, and Frank Harvey. Ann Arbor: University of Michigan Press, 176–91.