

New Approaches to Causal Analysis in Policy Research

Paul F. Steinberg
Department of Humanities and Social Sciences
Harvey Mudd College
paul_steinberg@hmc.edu

Paper prepared for the panel "Multi-Methods in Qualitative Research," Annual Convention of the American Political Science Association, Chicago, September 2-5, 2004.

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 whale populations to rebound; and that health education campaigns will reduce HIV transmission. As Pressman and Wildavsky (1973, xxi) observed a quarter century ago, "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 research, where they are used to diagnose problems, project future impacts of new regulations, and evaluate the effectiveness of - and assign responsibility for - past interventions. This paper focuses on one research practice of special importance to policy-oriented political scientists: the production of valid causal explanations through historical process tracing.

Causation means that the presence of a given variable changes the likelihood of a given outcome (see Pearl, 2000). Historical process tracing is the practice of analyzing causation by examining the streams of events mediating the relationship between antecedents and outcomes (George and McKeown, 1985; Bennett and George, 2001). The rationale for assessing causation through historical analysis derives from the fact that "Causal analysis is inherently sequence analysis" (Rueschemeyer et al., 1992, 4, quoted in Thelen, 1999, 390). Moreover, the outcomes of concern to policy researchers typically involve complex chains of events unfolding over time, which defy accurate characterization through regression analysis, with its more static snapshots of covariation. Given this, it is unsurprising that historical process tracing has played a central role in the canons of policy studies, where it has been used to shed light on the nature of decision-making (Allison, 1971; Tuchman, 1984), to evaluate implementation and effectiveness (Bardach, 1977;

Haas, Keohane, and Levy, 1993; Young, 1999), and to understand the evolution of institutions, policy processes, and social demands (Kingdon, 1984; Baumgartner and Jones, 1993; Grindle, 2000; Ostrom, 1990).¹

While few political scientists would dispute the importance of analyzing historical processes, there is little agreement on how to do it well and on the larger question of its role in producing verifiable propositions about cause and effect. Indeed, political scientists seeking rigorous approaches for assessing causation through historical methods face a recurring dilemma. Conventional correlational techniques are too blunt a tool for describing in meaningful detail the mechanisms that account for observed covariation between antecedents and outcomes. Yet few well developed alternatives exist that might combine the historian's craft with the social scientist's commitment to the systematic evaluation of causal claims. In recent years there has been a movement within political science to rectify this situation by developing new approaches to qualitative research.² Building on this effort, in this paper I propose several tools intended to refine the practice of historical process tracing as it applies to policy studies. My goal is to expand the menu of techniques for the collection, analysis, and presentation of information on causal pathways, with special emphasis on the challenge of producing valid causal explanations.

In the sections to follow I first discuss the special role of qualitative research in policy studies, joining the rising chorus of political scientists eager for new tools appropriate for small-N research settings. I then consider the value, as well as the challenges, of eliciting causal assessments from actors participating in policy processes and I recommend strategies for overcoming some of the more common sources of bias in his approach. This

¹ Each of these studies relies heavily on qualitative assessments of historical processes in order to assess cause-and-effect relationships, whether or not they use the term "process tracing."

² For an introduction, see the inaugural issue of the *Newsletter of the American Political Science Association Organized Section on Qualitative Methods*, Vol. 1(1), Spring 2003.

is followed by a discussion of latent causal pathways - those pathways to policy failures which may have been avoided due to the diligence of policy reformers, but nonetheless represent valuable and often unreported sources of information about causality and the prospects for success. I then turn to the practice of causal narrative, noting that the narrative form lends itself to strategic chronological expansion, which can be used to leverage additional information about the workings of causal mechanisms. In the final section I argue for greater attention to "mid-range methods" - consensus criteria for assessing the validity of specific categories of causal hypotheses.

Qualitative Research in Policy Studies

Beyond Correlation

Policy studies have long followed and inspired methodological debates in political science. This can be seen in the evolution of policy studies from the field's early emphasis on legal-descriptive analysis, through the focus on organization theory, systems research, game theory, operations research, econometrics, and interpretive approaches (for an overview see Friedman, 1987). In this context, the debate generated by the publication in 1994 of King, Keohane, and Verba's *Designing Social Inquiry* warrants careful attention by policy-oriented researchers. Arguing that all empirical social inquiry must conform to the tenets of valid statistical inference, these authors accuse those who eschew these rules of not only violating the requirements of correlational techniques, but of breaching the rules of sound science. Critics of King and colleagues raise the argument - to be reiterated here - that correlational analysis is too restrictive a conception of scientific practice, needlessly cordoning off a wide range of techniques that might be applied in the service of rigorous social inquiry (Brady and Collier, forthcoming). Scholars associated with the Consortium on Qualitative Research Methods have argued that while the world is not indeterminate - it is

knowable to some degree, and inter-subjectively verifiable propositions about causation are feasible - most social phenomena of interest to political scientists are multivariate, interacting, historically contingent, and probabilistic. The ontological underpinnings of correlational techniques are poorly suited to the study of these complex phenomena (Hall, 2003; Ragin, 1987).

This debate has special relevance for policy research, for at least two reasons. First, policy-oriented researchers must often evaluate causal effects in small-N settings. Policy research often focuses on small numbers of cases that show a unique departure from the norm - whether they are cautionary tales or exemplary accomplishments - and which therefore contain important lessons for the larger universe of policy practice. Peluso's (1992) study of teak management in Java, for example, shows in compelling detail how conservation policies may be used by authoritarian governments as a pretext for resource appropriation from marginalized rural populations. We need not establish a global statistical trend to take this and related cases seriously as a cautionary tale for those working to create national parks in the tropics.

Similarly, there is a high demand for insights into the rare and exemplary accomplishments of outliers such as the health prevention programs of the state of Ceará, Brazil (Tendler, 1997) or Indonesia's pollution reduction program (BAPEDAL and World Bank, 1995). Small-N studies are also important for evaluating pilot projects, which are typically implemented on a limited scale to allow innovators to experiment with new ideas while minimizing the cost of potential failure.³ Policymakers and others working in the public interest want to learn about the art of the possible, and the risk of the unthinkable, not

³ Statistical methods are useful in those small-scale policy experiments for which the results can be measured as individual effects (behavior change, customer satisfaction) in a large population. In contrast, understanding the exemplary land conservation efforts of one municipality among a hundred requires an intensive and non-correlational study of the outlier.

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 September 11th attacks.

There is a second reason why policy research requires tools that transcend the limitations of correlational analysis, and this concerns the role of causal mechanisms. Several authors have noted that regression analysis is of limited value in tracing causal mechanisms (Dessler, 1991; Tilly, 2001). Yet to design intelligent policy interventions, we need precise characterizations of the mechanisms through which posited causal variables exert their effects. Just as all public policies imply theories about cause and effect, different causal theories carry divergent implications for policy - accurate diagnosis is essential (see Bardach, 2000).

A case in point is international population policy. For years the observed correlation between rising national income and declining birthrates was interpreted, under demographic transition theory, to mean that economic growth was the surest path to mitigating spiraling population growth. These findings found fertile ground in the ideologically charged atmosphere of the 1974 World Population Conference in Bucharest, where developing country delegates wary of population policy argued that "Development is the best contraceptive" (United Nations, 1993; Finkle and Crane, 1985). Later analyses revealed, however, that there are many pathways to demographic transition (Cleland, 1993) and that a major proximate cause of reduced population growth is the availability of family planning programs - a variable that correlates with economic growth but is not wholly dependent on it (Bongaarts *et al.*, 1990). Bangladesh, for example, achieved significant reductions in population growth despite pervasive poverty by making large investments in outreach

programs for women.⁴ This more precise characterization of the mechanism linking economic growth and population stabilization produces starkly different policy prescriptions, with notable implications for the relative role of markets and states.

Process Tracing

The central challenge at this point in the methodological debate is to develop a suite of new techniques for qualitative analysis, to avoid a false choice between the well-developed canons of a methodological school inappropriate to the task at hand, and a free-for-all of case study methods based loosely on the practice of historians but lacking precise causal techniques. The most promising candidate for such an approach is process tracing, defined by George and McKeown (1985, 35-36) as studying the "process by which various initial conditions are translated into outcomes." In contrast to correlational approaches, process tracing evaluates "a stream of behavior through time... Any explanation of the processes at work in this case thus not only must explain the final outcome, but also must account for the stream of behavior." In essence, process tracing 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).

Bennett and George (1997) distinguish the practice of process tracing from conventional historical narrative as follows: "A process-tracing explanation differs from a historical narrative, as it requires converting a purely historical account that implies or asserts a causal sequence into an analytical explanation couched in theoretical variables that have been identified in the research design." They develop the concept in much

⁴ Bangladesh, Still Poor, Cuts Birthrates Sharply, *The New York Times*, September 13, 1994.

greater detail than did the earlier work by George and McKeown, laying the relevant groundwork from philosophy of science and describing different applications of process tracing in the development and testing of theories (Bennett and George, 1997, 2001).

Although significant strides have been made, process tracing is still in its infancy. It is best understood as a general category of research practice - a solidly constructed toolbox in need of specific instruments that researchers might employ in various combinations at the research design stage and during the practice of field research, when numerous decisions must be made about which paths to explore and how to do so. A few promising approaches have been advanced which, like Bennett and George's work, take advantage of the rich store of information produced by thick description and idiographic approaches and attempt to analyze this information more systematically. Recent work on counterfactual analysis has made important advances in systematizing the practice of imagining counterfactual scenarios in order to locate the causal importance of historical developments (Fearon, 1991; Tetlock and Belkin, 1996). In a different vein, Bates and colleagues (1998) use historical case studies to inform rational choice models. Building on these and related efforts (see Mahoney, 2000; Roberts, 1996), in the following sections I offer several additional strategies for assessing policy causation through process tracing.

The Role of Reported Causation

Process tracing in policy research often relies on interviews with actors who participate in and closely observe the process under investigation. Although newspapers sporadically report event-manifestations of ongoing processes, they rarely provide a contiguous account of the origins and impacts of events or the links among them. As a result interviews are often the best, and sometimes the only information source for detailed insights into causal mechanisms, even in case histories with extensive archival records (see

Allison, 1971). Notwithstanding the rich store of information available from interviews with these actors, an influential number of social scientists - particularly those of a more quantitative orientation - are ill at ease with the notion of assessing causal pathways on the basis of interviews. It is therefore worth critically examining the basis for this concern before moving on to other considerations.

Is reported causation by participants in policy processes a valid source of information in causal analysis? To answer this question, let us begin by considering that policy processes - from international diplomacy to local land use decisions - are driven by actors whose professional survival and personal satisfaction depend upon at least a rudimentary (and often extremely sophisticated) ability to assess the world as it is. The condition of purposive, reflective subjects offers social scientists a unique window on causation unavailable to our counterparts in the natural sciences. Our subjects wish to have causal effects, they go to great lengths to do so, and they typically reflect on - and are often held accountable for - the efficacy of their causal strategies.

Regardless of where scholars place themselves in agent-structure debates, nearly all would agree that human agency plays an important role (see Dessler, 1989). As Olson (1993, 16) points out, "the whole notion of intentionality and agency is grounded in the assumption that the agent can understand the ends to which his actions are directed and anticipate their consequences." Even if this assumption is not always true - indeed, even if it is frequently false - it must be true with substantial frequency for agency to play its widely acknowledged role in human affairs. The alternative would be to assume that people routinely achieve their desired ends by pure accident. Given the complexity of the social ends studied by political scientists - electoral strategies, policy reform, social mobilization, and the like - this seems highly implausible.

Many analysts are quick to point out the biased assessments produced by policy

actors - the self-aggrandized memoirs, strategic interview responses, lack of awareness of structural variables, and the like - as if these render these data less amenable to scientific inquiry. However every widely used scientific method carries serious limitations. The dependability of water quality sampling is compromised by variable rainfall patterns. Remote sensing data fail to distinguish among micro-level variations in plant cover. The fossil record is biased toward organisms with hard skeletal structures. Radiocarbon dating can give "a warped perspective of time" (Reimer, 2001), hedonic pricing methods fail to capture the value of non-consumptive goods, voting studies fail to reflect the intensity of individual preferences. None of these sources of bias or error provide justification for willfully ignoring the wealth of data they produce. Rather, clever ways must be devised to anticipate, attenuate, and report these limitations. The practice of eliciting reliable information from biased and error-prone sources is not only consistent with scientific practice - it is central to the craft.

The challenge is to devise strategies to deal with these inaccuracies. With this in mind, let us examine a source of bias so commonly encountered when using interviews and related techniques to trace causal pathways that we may label it "the fundamental bias of intentions." The fundamental bias of intentions has two components: First, during interviews and other self-reported assessments (newsletters, annual reports, personal memoirs) actors tend to report on their efforts and activities rather than their impacts. Second, as is widely recognized in structural critiques of agent-based models, these actors are likely to overlook and underreport the degree to which their causal impact is enabled or constrained by macro-level conditions. I will address each of these issues in turn. I argue that when attempting to accurately characterize a causal process, one should use an iterative approach along two axes: lateral iteration between adjacent links (dependent and independent variables) in a causal chain, and vertical iteration between macro-structural

causes and the specific configurations of events they shape.

Lateral Iteration

To assess the causal influence of variables in a chain of events leading to a policy outcome, one must distinguish between the activities and impact of the agents aspiring to have influence (see Steinberg, 1999). Failure to make this simple distinction has led to a profusion of ambiguous causal statements in the published literature. The type of statement that is the target of my critique includes: "The main effect of the project has been to increase public participation." Or, "Greenpeace affects global environmental politics by lobbying governments and by shaping public opinion." In each case it is unclear whether the phrase is describing activity or causation. Greenpeace may spend a lot of time on public opinion, but is public opinion shaped largely by Greenpeace? Participation may be the main effect of the project in question, but is that effect significant relative to the various forces shaping participation? The problem with this type of causal statement is that it is based entirely on observations of the posited independent variable. The crux of my argument is that after observing the effects of an actor on a particular outcome, the outcome itself must be studied more systematically to make a determination concerning the importance of the actor's efforts relative to the larger constellation of variables shaping the outcome. This, in turn, may lead the investigator back upstream to study these other causal precedents in greater detail. I refer to this as lateral iteration.

The problem of conflating activities and impacts is pervasive in assessments of the impact of international environmental actors and institutions. The rapidly evolving facts of global environmental politics beg explanation, and researchers are often in the position of chasing some interesting new phenomenon - an international treaty, an advocacy campaign, or an industry initiative - in an attempt to characterize its impact. Too often these

analyses neglect to study in any comprehensive way the intended target and the mechanisms and multiple causes that affect it. An example can be found in Keck and Sikkink's (1998) analysis of transnational activism. Noting that policy reform is one of the intended goals of international environmental campaigns, these authors conclude that policy change results from a "boomerang effect," whereby domestic advocates in developing countries enlist the help of foreigners to leverage changes in domestic policy. Yet in research on the causes of environmental policy reform in developing countries, the boomerang effect - and international campaigns more generally - are found to be of minor importance (Steinberg, 2001). The reason is that while policy is important to international campaigns, the reverse is not true. Campaign organizations expend a great deal of energy trying to affect policies, and these activities figure prominently in their promotional literature. But the investigator who "reads off" causation from these efforts will be misled. Campaigns are rare, narrowly focused, short-term events, and their impact is small relative to the decades-long process of policy reform.⁵

Conversely, actors may have a substantial impact on an outcome that is of relatively minor importance to them. Continuing with the example of policy change in the tropics, expatriate research ecologists expend little of their energy on matters of policy. It is a point of professional pride that they spend most of their time in remote field locations, and they will frequently focus the interviewer's attention on their dominant activity. But because they are among the few foreign environmental actors who remain in one tropical country over a

⁵ This raises the question as to why individuals or organizations would devote so much energy to matters on which they have relatively little influence. This may simply be a function of the limited size, duration, or leverage of the group relative to a challenging problem, and does not imply inefficient or irrational behavior. An organization may deem it worthwhile to affect outcomes even slightly, taking full advantage of its limited influence, if it is part of a larger effort or if its activities will produce small but tangible gains, such as saving one species out of 30 million or providing housing to a few low-income families in a society rife with poverty. Moreover, actors do not merely pursue strategies that will maximize their impact; their goals and methods are also shaped by their organizational history and ideology (Dalton, 1994).

period of decades - long enough to see through major policy developments - the small percentage of their time devoted to policy questions has an inordinate amount of influence.

This carries practical implications for the conduct of field research. In cases where there is a significant mismatch between activity and impact - a small proportion of actor effort driving an outcome, or a large proportion of effort producing a relatively minor impact - it is easy for researchers to be misled by the causal assessments provided by participants. Participants prefer to discuss the subset of their impacts that follow from their main interests and associated activities. Interview questions focused on a dependent variable that the actor shapes but does not care about are often perceived as annoying. The actor may feel that the interviewer is missing the point of what they do.

The notion that an investigator should have a clear understanding of the dependent variable in a study may seem so obvious as to not be worth mentioning. In the course of process tracing, however, a researcher must grapple with a diverse collection of social phenomena - from the origins of church-based social mobilization to the passage of civil rights legislation, for example. The different stages in the process - elite strategy, coalition bargaining, repertoires of repression and protest, lobbying, agenda setting - are governed by distinct forces and are often associated with diverse research traditions spanning several subfields, if not disciplines. For less studied historical sequences, it is difficult to anticipate during the research design stage many of the intervening variables one will encounter. Reading impact off of activity can be a tempting shortcut for this problem, but it should be resisted in favor of lateral iteration.

Vertical Iteration

One of the great advantages of investigating causal mechanisms is that researchers can move beyond arm's length correlational measures and gain a more nuanced

understanding of the chains of real world events that mediate the influence between variables observed to move in tandem. The worm's eye view, however, has its disadvantages. As noted in the previous section, causal process tracers typically encounter diverse and unanticipated causal variables during the course of their investigations and they may be unfamiliar with the more general phenomena of which their discoveries are examples. Moreover, as is often pointed out by those who focus on structural explanations, purposive agents and other participants in policy processes often overlook the larger "conditions of action" (Mahoney and Snyder, 1999; Dessler, 1989). One must therefore move vertically - looking for opportunities to understand field observations in light of broader patterns and their associated literatures, and using these literatures to produce new hypotheses and generate new lines of questioning.

A researcher may discover, for example, that a social security policy initiative failed to move through Ecuador's legislature for lack of multipartisan consensus. It behooves the researcher to make a vertical move, learning more about inter-party dynamics in that country, which can help discern whether this was an exceptional circumstance or the norm, and how the observed condition compares to circumstances in other countries where the researcher hopes to transfer lessons from the study in question. This approach differs from the "funnel strategy" described by Mahoney and Snyder (1999), because rather than moving unidirectionally from macro-structural to lower levels of analyses only after exhausting the explanatory power of the higher level, the iterative approach recognizes that one may become aware of the relevant structural questions only after exploring micro-level processes, and the insights gained by upward movement will produce new questions that can only be answered by returning to micro-level analysis.

The stretch the explanatory power of macro-structural versus proximate causes of an outcome are often debated as if they were rival explanations. More often than not the

rivalry in question is an artifact of researchers sorting themselves into distinct research traditions concerned with phenomena at different levels of generality. They may not in fact be rival hypotheses - which must be mutually exclusive - but simply different components of the set of conditions that are collectively sufficient to produce the outcome. Let us consider the example of the village of Ait Iktel, in the Atlas Mountains of Morocco, which has recently achieved dramatic improvements in public education and water and power services. According to a recent World Bank analysis (World Bank, 2003, 75-76), this outcome resulted from the entrepreneurial activities of a local NGO that mobilized community interest, building on the village's considerable social capital, and leveraged donor support from the Japanese Embassy. That 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 a result of two dynamic local leaders who had strong capabilities in external fundraising and project management. In this case a variety of conditions from macro- to micro-scales - political liberalization, decentralization, external resources, and a local NGO under entrepreneurial leadership - were all probably necessary to produce the observed outcome of improved social services.

This example, with its multifaceted and nested set of causal conditions, is typical of the phenomena studied by policy researchers and by political scientists generally. Is it possible in these situations to make meaningful statements regarding the relative causal strength of the component factors? After all, a great attraction of regression analysis is its careful delineation of the relative contributions of different independent variables. How might relative causal strength be operationalized in the context of small-N studies?

To answer this question we must recognize that causal outcomes are the result of sets of conditions. A cause-and-effect relationship is an emergent property of that set of interacting conditions. Just as it is meaningless to speak of the relative causal importance of flour and water in producing baking dough, the relative causal importance of an ingredient in a policy outcome cannot be specified in the abstract. The way out of this apparent dilemma is to tether the analysis to a social or analytic goal that can discern among component causes. In the context of policy research, three such goals are variation, attribution, and leverage. Correlational approaches focus on *variation* among cases. This is what allows regression analysis to discern each variable's relative importance, by removing from the analysis those variables that may be important causes in a more general sense (the Big Bang, the presence of functioning markets) but are constant across all cases. In contrast, *attribution* teases out the subset of causal conditions that can be attributed to moral agents. Thus of the many factors contributing to the September 11 terrorists attacks, the 911 Commission focused on the subset of causal conditions under the control of U. S. defense and intelligence agencies with a legal mandate to defend the country against attack. Under the attribution criterion, the failure to prevent an outcome is often treated as a cause of that outcome. *Leverage* focuses on those components most susceptible to manipulation. Thus the development of modern capitalism may have explanatory power as a cause of homelessness, but it is deemed irrelevant by the analyst searching for solutions, who might focus instead on housing costs.

For such leverage-oriented research, one approach is to rest the analytic lens on that part of the causal stream where fulfillment of the conditions present at that point would substantially increase the probability of producing the desired outcome "downstream" in the causal process (Steinberg, 2002). In the Ait Iktel case the lens might rest on "local social capital mobilized by an NGO with dynamic leadership and access to external resources."

The analysis then requires looking upstream to discern the variables (structural or otherwise) needed to arrive at that point in the causal stream, and focusing on those deemed most relevant for the purpose of leverage. In this case one might choose to downplay the fundraising skills of the NGO leaders if it is judged that there are other feasible ways to make resources available.

Leverage may be understood as the obverse of social structures, which are defined as those conditions not of our choosing and often beyond our control. From the above example, it is apparent that both concepts are at least partly in the eye of the beholder. Just as causation is a relational concept, the immutability of social structures may differ across actors. An assessment undertaken by a village entrepreneur might focus on fundraising skills as a necessary condition for building a new school, given the fixed rules of the game, whereas a policy reformer within the education ministry might view resource distribution channels as a relatively easy problem to fix, placing the analytic lens further upstream on the need for budget increases. The funnel approach described by Mahoney and Snyder (1999), in which one exhausts the explanatory potential of macro-structures before moving downward, may serve well the purpose of comparative historical analysis, helping to discern broad patterns; from the perspective of policy analysis, however, institutional change generally becomes more costly and difficult as one moves from micro- to macro-levels, and it may be preferable to move in precisely the opposite direction.

Latent Causation

Process tracing typically requires a significant commitment of time and resources, and thus there is a need for techniques that can help produce more causal information from a given analysis. One such strategy involves tracing not just the pathways that materialized, but those that would likely have come about were it not for the directed efforts

of policy actors committed to a particular path.

To appreciate this point, let us begin with the observation that there will always be more causes of policy failure than causes of success. This is so because a policy success is, by definition, faithful adherence to a complex set of causal events needed to conclude a treaty, to manage a national park, or to improve the health status of refugee populations. Any significant deviation from this particular chain of events (i.e., all other imaginable outcomes) constitutes a failure. Pressman and Wildavsky (1973), after counting no less than 70 clearance points along the path to successful implementation of a federal jobs program, conclude "It's amazing that federal programs work at all." Moreover, success generally occurs by design, whereas failure can occur by accident - everything from unanticipated consequences to inaccurate diagnoses, overlooked sources of opposition, corruption, turf battles, economic shocks, budget cuts, political turnover and countless other factors are sufficient to move a policy initiative onto a path leading to failure (Steinberg, 2003, 11-12). Failure has random events on its side. Indeed, public policies would not attract the considerable resource commitments required for their implementation if their purpose was to produce outcomes that are highly likely to result from chance alone.

As a consequence, any policy success implies averted failure. When tracing pathways, therefore, one can provide a richer causal explanation by revealing not just the manifest causes of success, but the latent causes - the mechanisms behind averted failure. This allows two things. First, it provides a better explanation of the connection between cause A and outcome B, because one is tracing not just the successful path but the unsuccessful ones and how they were averted. Studying policy success without studying averted failure is akin to studying peace without an understanding of conflict resolution. In both cases the "negative" outcome (war, or policy failures) is such a prevalent feature of society that its absence begs explanation.

Second, this approach produces information about latent, potential causes of failure that may express their deleterious effects when manifest causes of failure are resolved. For example, we might observe that wildlife officials in a poor tropical country are unable to halt illegal exploitation of endangered species for lack of funding. Inadequate budgets are the manifest cause of failure. If we resolve this problem, however, we might learn that a wildlife agency is rife with corruption - a latent cause of failure that did not have a chance to show itself because of the (manifest) lack of funds. An anti-corruption program is then implemented, only to reveal that officials lack proper enforcement training. Training is subsequently improved, and the initiative fails because of poor relations between enforcement officials and local communities - and so on. There is a long line of latent variables waiting in the queue.

Revealing averted failures and other latent pathways in a given case study can offer a richer set of information, including cautionary tales and effective mitigation strategies, to inform policy efforts elsewhere. One useful technique for tracing latent causal pathways is "revealing inertia" during interviews with key participants in policy processes (Steinberg, 2001). Process participants are prone to recount an event as if it arose as part of the natural and inevitable progression of events. As Tetlock and Belkin (1996, 15) point out, the cognitive psychology literature suggests that "Once people learn the outcome of an event, they not only perceive that outcome as more likely ex post than they did ex ante...they often fail to remember their exit ante assessment of what was and was not likely to happen." With the revealing inertia technique, as the respondent shares his or her narrative recollection of the stream of events producing a given outcome, the researcher imagines potential barriers and sources of inertia common to similar policy efforts and poses these as questions: "The Senate has many pressing matters, so how did your bill get to the front of the queue?" "Alliances among advocacy groups can be difficult to maintain - how did it

work?" "Did this unprecedented level of foreign assistance spawn jealousy or turf battles within the agency?" The hypothetical obstacles posed should derive from an understanding of problems frequently cited in the relevant literature, from information gathered in previous interviews, and from the process of vertical iteration described earlier. From the nature of the response it is usually obvious whether one has revealed or merely invented a source of inertia. The interviewer should monitor the level of enthusiasm or concern expressed, the detail and rapidity of the reply, and whenever possible verify the information in subsequent interviews.

The revealing inertia technique recognizes that the causal link between two variables in significant policy undertakings is not properly represented as the straight thin arrow common to causal diagrams. A more accurate symbolic representation would be a topographic map, in which the connection between origin and destination is fraught with steep climbs, harrowing cliffs, and sandpits. Considerable directed energy is needed to navigate this causal topography, overcoming sources of inertia and ushering initiatives toward a desired end point. Viewing policy causation in this light, researchers can use the revealing inertia technique to leverage more information about causal processes from a given case study.

Causal Narratives

Though the distinction between political scientists and historians is often characterized in terms of political scientists' preference for theoretical parsimony versus historians emphasis on descriptive richness, this distinction breaks down if we add explicit theoretical criteria for more or less richness or parsimony. This suggests two axes - richness versus parsimony and theoretical versus idiographic interest - along which process tracing may be situated in the information rich, theory-oriented quadrant. What are the

implications of a more theoretical orientation - and a specific interest in causal mechanisms - for the actual practice of gathering and reporting historical information? Are historians and political scientists crafting essentially similar case histories but using them for different purposes? Or should the craft itself differ in certain respects? Here I suggest one important way in which the historiography itself can and should differ, focusing on the construction and telling of historical narratives.

Every causal process can be described at a range of time scales corresponding to events embedded within larger events - for example, from "regime change" at a broad level, to the planning of the military coup that precipitated it, to negotiations among senior officers during the initial planning meeting, to the tactics of persuasion employed during these negotiations, and the rhetoric and nonverbal cues used to make those tactics effective. When constructing and reporting a causal story, process tracers must decide how much analytic effort and page space to devote to each component event. Narrative always involves strategic choices about how to compress and expand chronologies - even thick description is by necessity "lumpy" - and these choices are made according to the purpose of the teller. A skilled raconteur such as Mark Twain or J.R.R. Tolkien moves between brief mention of overarching events comprising the plot structure, and prolonged descriptions of the smallest details of a story, grabbing the listener's attention by drawing out the minutia of a comical situation or a frightening scene. Just as storytellers compress and expand narrative timelines for the purpose of entertainment, historians do the same to illuminate details deemed inherently interesting or telling, such as Joseph Ellis's (2001) account of the duel between Aaron Burr and Alexander Hamilton: "In addition to being a fascinating tale designed to catch your attention, it introduces themes that reverberate throughout all the stories to follow by serving as the exception that proves the rule. Here is the only occasion within the revolutionary generation when political differences ended in violence and death

rather than in ongoing argument" (Ellis, 2001, 18).

When the purpose of narrative is to assess causation, one should selectively expand those components of a chronology that demonstrate, refute, or otherwise carry important implications for purported causal mechanisms. Portraying events in their "extended form" has the effect of producing more information with which the reader can evaluate the validity of descriptive and causal claims. An example is found in Graham Allison's (1971) analysis of the Cuban missile crisis. In order to test the explanatory power of his bureaucratic politics model against alternative approaches, he strategically expands the chronology of events in order to decompose a policy decision - the U.S. decision to respond to Soviet missiles in Cuba with a naval blockade - into progressively smaller component processes, from presidential decision-making writ large, to inter-agency rivalries, to the alternative conceptual lenses that agency directors used in reaching their divergent conclusions, to word choice in classified correspondence among senior officials. King and colleagues (1994) mention the value of decomposing single cases into a series of events as part of their larger recommendation to leverage as much theoretically relevant information as possible from a given case study. But they do not explore the implications in full: Because of the embedded temporal scales within any historical event, narrative is especially well-suited to the practice of decomposing a case into smaller parts.

Mid-range Methods for Assessing Validity

What criteria are available for assessing the validity of causal narratives? Although it is unproductive to hold causal explanations to the standard of absolute certainty, reasonable certainty should be attainable. In statistical analyses, a common criterion is the 95 percent probability that a given correlation did not arise from chance alone. By what metrics might we judge the plausibility of causal claims based on historical process tracing?

Huntington (1991) observes that a satisfactory explanation is the place where the mind comes to rest. But this begs the question, *Whose mind?* Let us consider first the mind of the individual researcher. Roberts (1996) shows how historically-oriented researchers use miniature covering laws - well-known empirical regularities operating at the micro-level - to judge the plausibility of hypothesized causal mechanisms. For example, upon hearing that a mid-level official acted in response to a prime minister's phone call, the researcher - guided by the knowledge that such actions commonly produce the observed result - may deem that no further explanation is warranted. Hearing that a prime minister acted in response to a phone call from a mid-level official, however, the skeptical researcher will explore further to discern whether and how such a rarity came to pass. The strongest practitioners of historical analysis are those intimately familiar with a given historical context; by virtue of their immersion in the subject matter, they have accumulated an enormous stock of such micro-laws with which to judge the plausibility of competing claims.⁶

A second approach for assessing the validity of hypothesized causal processes - one that has received relatively little attention in qualitative social science - is to rely on groups of researchers familiar with particular types of causal mechanisms, who can produce research protocols, rules of thumb, and agreed-upon validity tests unique to that particular topic. Despite the retreat from a search for universal laws to greater emphasis on mid-range theory, there is a tendency to use a "universal method" approach to the design and assessment of research - asserting one common set of criteria for assessing the validity of any hypothesized causal pathway. Within the field of international relations, for example, Bennett (2003) observes that the self-conscious delineation of competing research

⁶ Roberts' larger point is that many such micro-laws exist. Where the Hempelian covering law approach fails is at the macro level, which by virtue of its aggregate nature and smaller number of observations typically defies law-like regularities. Causal analysis can therefore serve as a unifying theme between historiography and social science - but only if historians are more explicit about their use of micro-laws and if social scientists recognize that law-like regularities do occur, but at smaller scales.

programs has provided greater clarity but distracts attention from empirical, problem-based research on the causal mechanisms driving phenomena of interest to all parties. Bennett proposes "typology theory" as an alternative approach, focusing attention on recurrent categories of causal mechanisms that make up, in various combinations, the streams of events leading to specific and often unique historical outcomes. By focusing on the components of complex events, this approach recognizes the contingency of historical outcomes - a contingency that resists description by covering laws - without abandoning the search for recurrent patterns of social behavior.

To operationalize typology theory in policy research, we would do well to adopt a practice common to the natural sciences, which is for groups of research specialists to elaborate agreed-upon criteria for establishing causation with respect to specific subsets of phenomena - a practice I term "mid-range methods." Within the field of medical microbiology, an example can be found in the criteria elaborated by Nobel laureate Robert Koch for establishing the cause of an infectious disease. Koch argued that in order to prove that a given pathogen causes a given disease, one must be able to do the following: Isolate the organism from affected individuals. If one infects a healthy individual with that organism, that individual should contract the disease. And the investigator should in turn be able to recover the hypothesized agent of disease from the sick individual. In the field of conservation biology, researchers have crafted a rule-of-thumb for wildlife officials charged with preventing species extinction. Known as the 50/500 rule, it states that isolated populations need a genetically effective population of about 50 individuals for short term persistence, and 500 for long-term survival (Franklin, 1980; Soule, 1980). Whether in the form of a protocol, rule of thumb, or other guideline, mid-range methods are tools in the service of those tracing specific categories of causal processes.

One potential application of mid-range methods in policy research concerns criteria

for establishing that a given group of advocates had a significant causal impact on a policy agenda. The rudiments of such an approach are in place. Huberts (1989) proposes 20 premises and rules for assessing the impact of a social movement organization in a single decision-making process, including such criteria as intention, access to decision makers, short timelines between efforts and impacts, congruence between policy outcomes and advocate intentions, and continuity of advocate participants. Arts and Verschuren (1999) operationalize a triangulation approach based on advocates' self-assessment of their influence (ego-perception), assessments by the targets of their influence (alter-perception), and independent analysis by the researcher. Jordan (1998) uses a similar approach to assess NGO impact in the Brent Spar case, in which a Greenpeace campaign reversed Shell Oil's decision to dispose of a large oil production structure in the North Atlantic. It would be reasonable to add, as an additional criterion, the extent to which the decision or policy under investigation incorporates and explicitly refers to concepts developed by the advocates. Because concepts typically carry a unique vocabulary, their presence in a policy and their origins in a given advocacy group are often easy to trace. For example, the advocacy group Pesticide Action Network devised and ran a decade-long multinational campaign to ban what the organization named the "dirty dozen" - chemicals of renown toxicity and persistence in the environment. In 2001, the United Nations Environment Program oversaw the signing of the Stockholm Convention on Persisted Organic Pollutants, focusing on precisely the same dozen chemicals, referred to in United Nations press releases as the "dirty dozen." Similarly, Rochon (1998) traces the social origins and policy impacts of concepts such as sexual harassment.

Mid-range methods can help to advance the agenda that Peter Hall (2003) sets out when he calls for greater attention to the predictions that competing theories make about the processes linking variables. Breaking down complex processes into discrete typologies

of events, or "chain segments" representing constellations of action seen repeatedly throughout history - efforts to pass a bill, influence a legislator, improve industry compliance, revise land use practices, and so forth - one can use mid-range methods to assess the validity of the components of a larger causal narrative. As noted earlier, process tracing brings researchers into contact with diverse dependent variables. Some links in the causal chain were neither anticipated nor studied by the investigator prior to the commencement of field research and may be governed by mechanisms beyond the investigator's area of expertise. To the extent that a menu of mid-range methods is available, process tracers will be able to not only situate the discovered variable in the context of insights from the relevant literature, but study it anew with the benefit of methodologies specifically honed for that purpose.

Conclusions

King, Keohane, and Verba (1994) make a compelling argument that science should not be considered synonymous with quantitative analysis; there is no reason for qualitative research to be any less scientific, in the sense of producing rigorous descriptive and causal explanations. We must now go one step further and erase the false assertion that rigorous qualitative research is synonymous with correlational analysis. Process tracing presents a potentially powerful alternative to correlational analysis, but it remains a general category of practice in need of specific tools that researchers can employ in various combinations according to the purposes and pathways of a particular analysis. We must not only critically re-evaluate the ontological underpinnings of correlational analysis, as Peter Hall, Charles Ragin, and others have done so effectively, but operationalize this new insight with process tracing techniques that become a regular part of research practice. Ultimately the claims of the "New Qualitative Methods" school will be judged by the degree to which the new

insights produce better research. In order to meet this test we must not merely provide a more sophisticated justification for the existing practices of historically-oriented researchers, but offer novel approaches to improve the craft.

This paper is one attempt to contribute toward this larger goal by devising new concepts and approaches for establishing causation in small-N policy studies. I have argued that interview subjects represent a rich resource for causal assessment, notwithstanding the associated challenges. Rather than discard the data, we must identify innovative methods for calibrating the biases and errors introduced by self-reported assessments of causation. Among these challenges is the need to recognize the fundamental bias of intentions, which leads actors to report activity rather than impact and to overlook structural variables bearing on policy outcomes. I suggest that this problem can be overcome by iterating laterally (between dependent and independent variables in a process stream) and vertically - examining the larger forces that shape intermediate dependent variables discovered along the way. To discern the importance of diverse antecedents operating at different scales, the concept of relative causal strength must be tethered to specific explanatory goals such as variation, attribution, and leverage. I then discussed the value of studying latent causal pathways and described the revealing inertia technique for doing so. I further argued that process tracing is uniquely suited to expanding the set of information bearing on causal claims through the strategic chronological expansion of historical narratives. To help evaluate the relative merits of causal assessments, I argue that we must develop mid-range methodologies in which experts come to agreement on criteria for assessing causation within distinct types of causal processes. As the policy examples presented throughout this paper should make clear, improvements in causal assessment hold not only the promise of better scholarship but of more intelligent approaches to solving social problems.

References

Allison, Graham T., *Essence of Decision: Explaining the Cuban Missile Crisis*, Little, Brown & Co., Boston, 1971.

Arts, Bas and Piet Verschuren (1999) Assessing Political Influence in Complex Decision-making: An Instrument Based on Triangulation, *International Political Science Review* 28(4):411-424.

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

Bardach, Eugene, *The Implementation Game: What Happens After a Bill Becomes A Law*, MIT Press, Cambridge, 1977.

Bardach, Eugene, *A Practical Guide for Policy Analysis: The Eightfold Path to More Effective Problem Solving*, Chatham House Publishers, New York, 2000.

Bates, Robert H. et al., *Analytic Narratives*, Princeton University Press, Princeton, NJ, 1998.

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

Bennett, Andrew, *Causal Inference in Case Studies: from Mill's Methods to Causal Mechanisms*, paper presented at the American Political Science Annual Convention, Atlanta, Georgia, 2-5 September, 1999.

Bennett, Andrew, *Beyond Hempel and Back to Hume: Causal Mechanisms and Causal Explanation*, paper presented at the American Political Science Association Annual Convention, Philadelphia, August 28, 2003.

Bennett, Andrew and Alexander George, Case Studies and Process Tracing in History and Political Science: Similar Strokes for Different Foci, in Colin and Miriam F. Elman (eds.), *Bridges and Boundaries: Historians, Political Scientists, and the Study of International Relations*, MIT Press, 2001.

Bongaarts, John, W. Parker Mauldin, and James F. Phillips (1990) The Demographic Impact of Family Planning Programs, *Studies in Family Planning* 21(6):299-310.

Brady, Henry E. and David Collier (eds.), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Roman & Littlefield, forthcoming.

Cleland, John, *Different Pathways to Demographic Transition*, paper presented at the Population Summit, New Delhi, October 1993.

Dalton, Russell J., *The Green Rainbow: Environmental Groups in Western Europe*, Yale University Press, New Haven, CT, 1994.

Dessler, David (1989) What's at Stake in the Agent-Structure Debate?, *International*

Organization 43 (3):442-355.

Dessler, David (1991) Beyond Correlations: Toward a Causal Theory of War, *International Studies Quarterly* 35:337-355.

Ellis, Joseph J., *Founding Brothers: the Revolutionary Generation*, Alfred A. Knopf Publishers, 2000.

Fearon, James D. (1991) Counterfactuals and Hypothesis Testing in Political Science, *World Politics* 43(2):169-195.

Finkle, Jason L. and Barbara B. Crane (1985) Ideology and Politics at Mexico City: the United States and the 1984 International Conference on Population, *Population and Development Review* 11(1):1-28.

Franklin, I. R., Evolutionary Change in Small Populations, in Soule, M. E., and B. A. Wilcox (eds.), *Conservation Biology: An Evolutionary-ecological Perspective*, Sinauer Assoc., Sunderland, MA, 1980.

Friedman, John, *Planning in the Public Domain: From Knowledge to Action*, Princeton University Press, Princeton, 1987.

George, Alexander L. and Timothy J. McKeown, Case Studies and Theories of Organizational Decision Making, pp. 21-58 in Robert Coulam and Richard Smith (eds.), *Advances in Information Processing in Organizations*, (JAI Press), vol. 2, 1985.

Grindle, Merilee S., *Audacious Reforms: Institutional Invention and Democracy in Latin America*, Johns Hopkins University Press, Baltimore, 2000.

Haas, Peter M., Robert O. Keohane, and Marc A. Levy (eds.), *Institutions for the Earth: Sources of Effective International Environmental Protection*, The MIT Press, Cambridge, MA, 1993.

Hall, Peter A., Aligning Ontology and Methodology in Comparative Research, in James Mahoney and Dietrich Rueschemeyer (eds.), *Comparative Historical Analysis in the Social Sciences*, Cambridge University Press, 2003.

Huberts, L. (1989) The Influence of Social Movements on Government Policy, *International Social Movement Research* 2:395-426.

Huntington, Samuel P., *The Third Wave: Democratization in the Late Twentieth Century*, University of Oklahoma Press, 1991.

Jordan, Grant (1998) Indirect Causes and Effects in Policy Change: The Brent Spar Case, *Public Administration* 76(4):713-__.

Keck, Margaret E. and Kathryn Sikkink, *Activists Beyond Borders: Advocacy Networks in International Politics*, Cornell University Press, Ithaca, NY, 1998.

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

Kingdon, John W., *Agendas, Alternatives, and Public Policies*, Little, Brown & Co., Boston, 1984.

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

Mahoney, James and Richard Snyder (1999) Rethinking Agency and Structure in the Study of Regime Change, *Studies in Comparative International Development* 34(2):3-32.

Olson, Richard, *The Emergence of the Social Sciences, 1641-1792*, Macmillan Publ., New York, 1993.

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

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

Peluso, Nancy Lee, *Rich Forests, Poor People: Resource Control and Resistance in Java*, University of California Press, Berkeley, CA, 1992.

Pressman, Jeffrey L., and Aaron Wildavsky, *Implementation: How Great Expectations In Washington Are Dashed In Oakland*, University of California Press, Berkeley, 2nd edition, 1973.

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

Reimer, Paula J. (2001) A New Twist in the Radiocarbon Tale, *Science* 294: 2494-2495.

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

Rochon, Thomas, *Culture Moves: Ideas, Activism, and Changing Values*, Princeton University Press, Princeton, NJ, 1998.

Rueschemeyer, D., E. H. Stephens, and J. P. Stephens, *Capitalist Development and Democracy*, University of Chicago Press, Chicago, 1992.

Soule, M. E., Thresholds for Survival: Maintaining Fitness and Evolutionary Potential, in Soule, M. E., and B. A. Wilcox (eds.), *Conservation Biology: An Evolutionary-ecological Perspective*, Sinauer Assoc., Sunderland, MA, 1980.

Steinberg, Paul F. (2003) Understanding Policy Change in Developing Countries: The Spheres of Influence Framework, *Global Environmental Politics* 3(1):11-32.

Steinberg, Paul F., *Environmental Foreign Policy in Developing Countries: A Capacity-Building Approach*, paper delivered at the International Studies Association Annual Convention, New Orleans, March 24-27, 2002.

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

Steinberg, Paul F., *The Measure of Success: Some Methodological Considerations for Research on Institutional Performance*, 5th Annual Colloquium on Environmental Law & Institutions, Duke University, April 27-28, 2000.

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

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

Thelen, Kathleen (1999) Historical Institutionalism in Comparative Politics, *Annual Review of Political Science* 2:369-404.

Tilley, Charles (2001) Mechanisms in Political Processes, *Annual Review of Political Science* 4:21-41.

Tuchman, Barbara, *The March of Folly from Troy to Vietnam*, Knopf, New York, 1984.

United Nations Secretariat, Population Division, Department of Economic and Social Information Policy Analysis, *Evolution of Population Policy Since 1984: A Global Perspective*, Population Policies and Programmes United Nations, New York, 1993, pp. 27-41.

World Bank, *World Development Report 2003*, Oxford University Press, 2003.

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