New England Journal of Political Science

Volume 2 | Number 1 Article 3

9-1-2006

Empirical Social Inquiry and Models of Causal Inference

David Dahua Yang

Empirical Social Inquiry and Models of Causal Inference

David Dahua Yang

Abstract

This essay examines several alternative theories of causality from the philosophy of science literature and considers their implications for methods of empirical social inquiry. In particular, I argue that the epistemology of counterfactual causality is not the only logic of causal inference in social inquiry, and those different methods of research appeal to different models of causal inference. As these models are often philosophically inter-dependent, a more eclectic understanding of causation in empirical research may afford greater methodological versatility and provide a more complete understanding of causality. Some common statistical critiques of small-N research are then considered from the perspective of mechanistic causal theories, and alternative strategies of strengthening causal arguments in small-N research are discussed.

> We political scientists know on some level that a true and complete explanation of the things that fascinate us would be impossibly complex, but we willfully ignore this disturbing fact and persist in our research. We are a community of eccentrics who share the delusion that politics is simpler than it appears. Although I would be as delighted as any other political scientist to discover

> simple, elegant and powerful explanations, I think the common sense of the layperson is correct. We must presume that politics is extremely complex, and the burden of proof rests on those who claim that it is not.—*Michael Coppedge*

Introduction

The logic of causal inference is a cornerstone issue in empirical social scientific research. In the opening pages of their influential work *Designing Social Inquiry*, Gary King, Robert Keohane, and Sidney Verba argue that a common logic of causal inference underlies both quantitative and qualitative methods, an assertion central to the counsel they then dispense to the qualitative researcher in the remainder of their book. Not surprisingly their unabashed enthusiasm for the "statistical worldview" has proven controversial. Much of the subsequent methodological debate has centered on this fundamental assumption of logical homogeneity in empirical social inquiries.

In this essay I will explore the relationship between methods of empirical social research and the theoretical models of causal inference that underlie these methods. Before I begin, I should acknowledge that causal inference is not the only worthwhile aspect of the empirical social scientific enterprise. In the words of Dietrich Rueschemeyer (2002), "tested empirical hypotheses and persuasive causal inference are not the only worthwhile products of social science research." Other tasks of equal importance include the identification of universal and quite

general problems, as well as the development of theoretical frameworks offering useful conceptualizations (Rueshemeyer 2002, 305-336). In fact a plausible argument can be made that since social science concerns human actions, social explanations should explore reasons for human action, but such reasons are not causes and should instead be understood in terms of meanings for social actors (e.g. Taylor 1977). As Henry Brady (2002a) observed social inquiry involves and depends on a broad spectrum of scholarly endeavors each contributing to the store of human knowledge in its own way, and it is not obvious why causal explanations should be automatically privileged over other types of explanations that are descriptive in nature. Even within the enterprise of causal inference, it does not necessarily follow that a stringent logic of inference should regulate all aspects of the research endeavor.

The process of theory discovery is quite distinct from the process of theory verification, and a plausible case can be made that contemporary social science is starving more of an effective albeit heuristic methodology for theory construction than of a rigorous methodology for theory testing (McKeown, 2002). As we shall see, methodological discussions of causal inference are likely to be the most productive when properly located within this larger context.

Alternative Theories of Causal Inference

Among statisticians the best-known theory of causal inference is the Neyman-Rubin-Holland (NRH) Theory that arose out of the experimental

tradition. The authors of *DSI* placed themselves squarely within this tradition and similarly adopted a counterfactual definition of causality (King, Keohane, Verba 1994, 76-78). But as Henry Brady (2002b) pointed out in his excellent exposition on the competing models of causal inference, the counterfactual theory that underlies the NRH approach is but one of several alternative theories in an ongoing philosophical debate. A brief overview of these alternatives will prepare the grounds for our discussion.

On a philosophical level questions about causality can be raised in three areas – psycholinguistic, ontological and epistemological. For our purposes as practical researchers we are mostly concerned with the epistemological question. That is to say, we are mostly concerned with the *discovery* of causal relations. On an epistemological level, Brady identified four major theories of causality.[1] These are:

> 1) Neo-Humean Regularity Theory, in which the definition of causality is based on observation of constant conjunction and correlation, with causal direction established by temporal precedence;
>
> 2) Counterfactual Theory, in which C is considered a cause for E if the statements "If C then E" and "If Not C then Not E" are true in

[1] Brady did not address the probabilistic theory of causation. The four major theories are deterministic – in a LaPlacian sense. Although the statistical approach is often associated with a probabilistic understanding of causation, statisticians within the NRH tradition (which is to say most) are in fact LaPlacian determinists, as A.P. Dawid pointed out in his rejoinder to critiques of his 2000 JASA article.

otherwise similar worlds, with causal direction established by the truth of the statement "If Not E then C may still occur";

3) Manipulation Theory, in which causality is established by a recipe of action that regularly produces the effect from the cause;

4) Mechanistic (or Process-based) Theory, in which causality is established by the discovery of mechanisms that lead from the cause to the effect, with the causal direction established by the operation of the mechanism.

Of the four theories, the limitations of the neo-Humean approach are well known. Even with refinements such as the INUS (Insufficient but Necessary part of a condition which is itself Unnecessary but exclusively Sufficient) definition of a cause, such an approach to causation cannot logically distinguish causation from correlation.[2] Likewise the naïve version of the counterfactual approach is confounded by spurious correlations and the problem of causal direction. The addition of a 3rd condition ("Even if the effect did not occur, the cause may still occur") can establish causal direction in some cases, but is of little value when the causes are independently sufficient because then they would never occur without their effects. The manipulation theory underlying the experimental approach solves the problems of causal direction and spurious correlation, but it can still be confounded by the problem of preemption (when redundant

[2] Indeed, as McKeown (2002) pointed out, neo-Humeans such as Carl Hempel wish to dispense with causation altogether, although such a position seemingly rebels against common sense.

sufficient causes are present but only some of which are manipulated), nor does it offer a causal explanation beyond the demonstration of causal effects.

In addition, both counterfactual and manipulation theories have come under attack by critics who insist that the social world is a densely interconnected system where changes to one part of the system would ripple through the rest, rendering the concept of an "otherwise identical" world or the isolated manipulation of causal variables logically impossible, leaving aside the fact that direct manipulation is seldom possible for the social scientist. In Brady's views the least problematic of the causal theories appears to be the Mechanistic Theory. Causation is established by the presence of causal processes between causes and effects. Observed regularities are explicated in terms of lower level processes and the mechanisms invoked vary from field to field and from time to time. Furthermore, the mechanistic approach provides a satisfactory solution the pairing problem (the pairing of particular causes and effects) that plagues the neo-Humean and counterfactual theories.

Although counterfactual causality is generally the norm in statistics, it should be noted that even among statisticians counterfactual theories are by no means universally embraced. Proponents of probabilistic causal theories assert that counterfactual arguments (and the associated assumptions required for statistical simulation of the counterfactual) are inherently "metaphysical" and "un-testable", and the counterfactual approach is therefore rather less "scientific"

than commonly believed. (e.g. Dawid, 2000; Shafer, 2001) I am not equipped to comment on this debate, except to note that a probabilistic understanding of causation such as that proposed by Sprites, Glymour and Scheines (2000) would require vastly different assumptions in statistical causal inference, although once again critics charge that the assumptions can seldom be met.[3] At any rate, it is far from obvious why counterfactual causality should be enshrined over alternative approaches in the social sciences.

Proponents of the NRH approach often insist that the counterfactual definition of causality is logically prior to the identification of causal mechanisms, and that the mechanistic approach would quickly lead to infinite regress as the identification of each causal linkage would require us to identify yet another set of causal mechanisms explaining that linkage (King, Keohane, Verba 1994, 85-87). Regarding the ontological priority of the counterfactual definition their assertion may contain a certain amount of truth. Stuart Glennan, a leading proponent of the mechanistic school, offered the following definition of causal mechanisms:

[3] Briefly stated, probabilistic causation holds that A causes B if and only if P(B|A & T) > P(B|~A&T) for all test condition T. In words, A causes B if the presence of A increases the probability of B occurring and no other factor can account for the increase. For a highly readable introduction to the topic, see Hitchcock 2002).

> A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations (Glennan 2000, 6).

For Glennan, causal mechanisms have relatively stable arrangement of parts and can manifest themselves at multiple locations in space-time. In this sense causal mechanisms are general. Furthermore, mechanisms consist of a number of parts displaying relatively stable properties in the absence of interventions. Given a particular decomposition of parts, "interaction" between parts is "a causal notion that must be understood in terms of the truth of certain counterfactuals," while "direct, invariant, change-relating generalizations" relate closely to the manipulation theory of causation as well as the neo-Humean "laws of succession".[4] In other words, each chain of mechanistic explanation must ultimately be subsumed within an appropriate covering law derived from observed regularities.

But as others have noted, Glennan's major concern was a methodological solution to causal inference for the practicing scientist. As such, the presumed *ontological* priority of the counterfactual definition does not necessarily demand a counterfactual *epistemology* as well. The distinction is particularly telling with

[4] By Glennan's definition, a "change-relating generalization" describes a relationship between two variables in which an intervention that changes one variable will bring about a change in another variable.

regards to the concern over infinite regress. As Glennan explained causal mechanisms are hierarchical complex systems, in the sense that parts of mechanisms may themselves be complex mechanisms that can be decomposed into further parts. However, the chain of explanation does not go on *ad infinitum*. Although all explanations must ultimately terminate at the bedrock of fundamental physical laws, the level of explanation of interest to the researcher typically "bottoms-out" rather more quickly. As Brady observed, "molecular biologists do not seek quantum mechanical explanations and social scientists do not seek chemical explanations of the phenomena they study" (Brady 2002b, 31).

The fact of the matter is that a good deal of research in the social sciences is informed at least implicitly by the mechanistic approach to causation. And, as noted by Larry Bartels (2002), it is neither feasible nor particularly fruitful to attempt to force all empirical research into the "Procrustean bed" of statistical counterfactual inference underlying the NRH tradition. When comparative scholars differentiate between "data set observations" and "causal process observations" (Collier, Seawright & Munck, 2002), they differentiate between counterfactual and mechanistic notions of causality. When they appeal to process tracing, "thick analysis" (Collier, Brady & Seawright, 2002) or "within-case control" (Munck, 2002), they appeal to the inferential power of causal process analysis and the logic of causal mechanisms. Critics have often pointed to the limitations of nominal comparative methods such as the Millian methods of

agreement and difference in small-N studies. However, I suspect in most cases these methods merely serve as heuristics for theory generation, while the true persuasive force of qualitative causal inference derives from the explanatory powers of mechanistic causality.

Combining Methods of Causal Inference

While on an ontological level these alternative theories of causality are often seen as competing or even contradictory, Brady argues that social scientists would do well to embrace them all. In fact, he believes that "a really good causal inference should satisfy the requirements of all four theories" (Brady 2002b, 2-4). Philosophers may be appalled, but as a practical matter I find his advice eminently sensible. As political scientists most of us are hardly equipped to adjudicate between the competing theories of causality, yet we can hardly afford to refrain from making causal inferences while the philosophers settle their debates. Indeed, given our current state of knowledge analytic paralysis would likely result if we were to insist upon an absolutist adherence to one model of causality or another. What we need, therefore, is a pragmatist, eclectic approach of the sort proposed by Rudra Sil and Peter Katzenstein (2005); an approach which—in the words of William James—allows us to sidestep "metaphysical disputes that otherwise might be interminable", and to instead "try and interpret

each notion by tracing its respective consequences" in concrete situations (Cited in Sil & Katzenstein, 2005; p.12). Empirical social inquiry is likely to be far more productive if we accede to a problem-driven view of scholarship, creatively matching and combining different ideas and theories according to their practical utility in different circumstances rather than confining ourselves dogmatically to any single (and much disputed) framework. In terms of concrete strategies for research this perspective recommends a "nested" approach to causal inference, which is also epistemologically well grounded in the overlap and interdependency between disparate theories of causality.

The truth may be that a complete understanding of causality cannot be derived from causal mechanisms or counterfactual/experimental causal effects alone. As Andrew Bennett and Alexander George (1997) observed, "a variable cannot have a causal effect on an outcome unless there is an underlying causal mechanism, and it makes no sense to define any entity as a causal mechanism if it has no causal effect." Because "causality involves both causal effects and causal mechanisms and its study requires a diversity of methods," meaningful progress in the social scientific enterprise is most likely when efforts are made to conjoin the quantitative assessment of causal effects with the qualitative assessment of causal mechanisms. Such a conjunction must transcend simplistic notions such as a division of labor between rigorous statistical hypothesis-testing and pre-

statistical qualitative hypothesis generation. It must leverage the inferential strengths of each method to redress the deficiencies inherent in the other.

It is important to bear in mind that both quantitative and qualitative methods are fraught with pitfalls and limitations. Process tracing alone has limited ability to estimate causal effects or the relative weight of multiple causal factors. Furthermore, the logic of mechanistic causal inference demands that each chain of explanation be terminated in a reliable covering law, yet off-the-self covering laws are scarce in the social sciences and the formulation of new laws is an essentially statistical endeavor. On the other hand statistical methods are built on a set of fundamental assumptions (SUTVA),[5] conditional independence of assignment and outcome) which are simply untestable in observational studies. Yet the validity of these assumptions are of profound consequences – If conditional independence does not hold, then regression results are likely to be biased; if SUTVA does not hold, then it may not even be possible to interpret the results (Brady 2002b, 43-49). Even when both conditions hold two difficulties are still inevitable: Firstly, correlation without the support of mechanistic explications may not satisfy our commonsensical notions of causation; and secondly, the estimated coefficients often do not correspond to counterfactuals of

[5] Stable-Unit-Treatment-Value-Assumption, which holds that the outcome for a particular case does not depend on what happens to the other cases or which of the supposedly identical treatments the unit receives.

substantive interest (See Sekhon 2002).[6] Then there are a number of familiar practical problems associated with each method. These include the small-N generalizability problem that plagues in-depth case studies due to their logistical demands as well as the measurement validity problem that plagues quantitative studies due to the difficulty of operationalizing concepts across many cases.

In light of these considerations, I find the growing interest among political scientists in nested analysis encouraging indeed. Method nesting cannot eliminate all difficulties, but an appropriate understanding of the underlying models of causal inference will no doubt help the researcher maximize leverage in combining methods. If we embrace a pragmatist and eclectic approach toward the logic of causal inference, there is no reason to insist upon any particular division of labor in theory construction. Depending on the model of causality invoked, an investigator may employ either qualitative or quantitative methods in theory generation and validation.[7]

---

[6] Sekhon gave the example of a statistical study of the relationship between race and uncounted ballots. "Before any regression is estimated, we know that if we measure enough variables well, the race variable itself should be insignificant. But in a world where being black is highly correlated with socioeconomic variables, it is not clear what we learn about the causality of ballot problems from a showing that the race coefficient itself can be made insignificant" (Sekhon 2002, 24).

[7] For some helpful insights on the methodology of nested analysis, see Bennett 2002; and Lieberman 2002.

Thus, a researcher may employ statistical methods to generate hypotheses, identify causal puzzles, or single out explanations of exceptional promise, but appealing to a mechanistic model of causation he can also rely on mechanism analysis to test hypotheses and establish causality. This does not imply that statistical methods would play no role in mechanistic causal inference, however. For one, mechanisms alone do not present a complete picture of causality and measurements of causal effects must be obtained through statistical means. But more fundamentally, the scholar invoking causal mechanisms may still have to rely on statistical methods to establish the underlying patterns of regularity, or he should at least be explicit about the untested assumptions built into the chain of causal explanations, which may then be subject to statistical verification later. For example, in the literature on democratic transitions it has often been asserted that the spread of liberal-democratic values is a key mechanism undermining the stability of authoritarian regimes as democratic-minded citizens actively oppose or at least withdraw support from their authoritarian government. While the argument may seem intuitive, its validity cannot be established without statistical investigation of the relationship between normative political values on the one hand and actual patterns of political behavior on the other. If political behavior is found to be driven primarily by instrumental considerations of material costs and benefits, then the mechanism linking democratic values and

declining authoritarian support may be spurious or at least seriously mis-specified.

At the same time, qualitative methods can be more than descriptive measurement devices or motivating heuristics for the quantitative researcher. As Brady (2002b) remarked independence of assignment and treatment is difficult enough to attain in a randomized experimental design, and the investigator who wishes to attain conditional independence in a "quasi-experimental" observational study faces daunting challenges indeed. She may have no way of ensuring conditional independence, although she may appeal to mechanism analysis to justify the otherwise "metaphysical" assumption of SUTVA, and rely on process tracing to establish causal direction and rule out spurious correlation. Consider, for instance, a study on protest movements in a newly-democratized country, in which the author argued that, since the former authoritarian regime saw the urban middle class as its primary base of support, it was far less likely to respond coercively to those movements which involved significant middle class participation. The author then demonstrated a statistically significant relationship between middle class participation and the absence of coercive responses.

Leaving aside the obviously thorny issue of SUTVA for the moment, upon more careful consideration the causal relationship involved may be far less straightforward than it first appears. Did the presence of middle class groups

deter the use of coercive measures, or did middle class groups only make their voices heard on issues which were not likely to invoke coercive responses to begin with? To what extent could the phenomenon be attributed to the fact that middle class groups were more likely to participate in protest movements during periods of political loosening, when the authorities were more restrained in the use of coercion? While these and similar considerations can be controlled for in a statistical model, oftentimes their very identification requires careful process-tracing through in-depth participant interviews and painstaking archival research. When the appropriate statistical indicators are not readily available, proper mechanistic explication becomes even more crucial. In this sense, causal process analysis can be an integral component of statistical research built on a basis of counterfactual causality.

Statistical Critiques of Small-N Causal Inference

With some basic understanding of the alternative models of causal inference in mind, we now turn our attention to a subject of much passionate debate in methodological discussions—namely, the contribution of small-N research in the process of theory construction. Although even critics generally allow that case studies are useful for identifying potential explanatory factors and generating hypotheses, the ability of small-N research to generate reliable causal inference has been a major point of contention. We begin with a brief overview of the critiques of small-N methods from a statistical perspective.

The two issues receiving the most attention in recent years are causal indeterminism and the closely related issue of selection bias. While many scholars have commented on the presumed dearth of logical and inferential constraints in small-N research, (e.g. Lijphart, 1971; Most & Starr, 1984; Achen & Snidal, 1989) probably the best-known formulation of the problem is given by King et al. in *Designing Social Inquiry*. A research design is said to be indeterminate when the number of observations is less than the number of variables under investigation. According to the statistical logic of NRH theory, in such a study not enough observations are available to test counterfactual hypotheses and consequently causal effects cannot be estimated. The problem is considered one of inferential logic, and not merely an artifact of linear algebra.

In addition, small-N research is said to be especially vulnerable to the selection bias problem. In an oft-assigned article Barbara Geddes (1990) argued that many seemingly plausible causal relationships suggested by political scientists – including most famously the relationship between foreign threats and social revolutions identified by Skocpol (1979)–disappear if more cases are taken into consideration.[8] Although Geddes was concerned primarily with selection on the dependent variable, her examples seem to suggest that selection bias is

[8] However, Skocpol and her defenders have argued that the Latin American cases added by Geddes do not belong in Skocpol's original domain of investigation, namely wealthy, politically ambitious agrarian states.

probable in small-N studies whether or not the investigator explicitly selected on the dependent variable. (Skocpol did include non-revolutionary cases in her work, and Hirschman was merely unfortunate in the timing of his article.) Furthermore, selection bias may be an issue even for the small-N researcher who does not care about generalizing her findings. According to David Collier and Jim Mahoney (1996), the true magnitude of certain causal effects may become apparent only when the entire universe of cases is considered. The researcher who focuses on a few cases risks being misled.

The methodology of small-N comparative studies has come under criticism as well. Several researchers (e.g. Lieberson, 1992; Sekhon, 2002) have argued that the Millian methods of agreement and difference often employed by qualitative scholars are especially inappropriate for small-N studies, since they do not allow for imperfectly regular causal patterns,[9] interaction effects, measurement errors or even the presence of common causes. Lieberson in particular constructed a well-known example on drunk driving demonstrating

[9] Although the issue is often referred to as "probabilistic causation", a more accurate description is "imperfectly regular causal patterns". Counterfactual causation is ontologically deterministic, and quite different from, say, the notion of probabilistic causation as proposed by Sprites, Glymour & Scheines. However deterministic causation can *appear* probabilistic if the causal model is imperfectly specified. See Brady 2002, 14-15.

how Millian methods in small-N studies can easily produce misleading causal inferences.

Following the same logic, Jim Mahoney (1999b) counseled that nominal small-N comparison and the methods of agreement and difference are only useful for *falsifying* theories—in fact, only *deterministic* theories where a single deviation is sufficient for falsification.

It should be recognized that not all small-N comparative studies aspire to generate causal explanations. As Skocpol and Somers (1980) pointed out, macro-causal theory construction is only one of three common objectives in comparative research. Other objectives include the parallel demonstration of theory (where cases are used to demonstrate the usefulness of a theory) and the contrast of contexts (where cases are used to illustrate differences between purportedly similar situations). One may argue that small-N studies designed primarily for theory testing and refinement should not be evaluated by the same standards for causal inference. However, some critics maintained that due to problems such as selection bias the small-N approach is too weak even to "provide honest tests of theories." (e.g. Achen & Snidal, 1989)

To counter the problem of causal indeterminism qualitative researchers are often advised to take the problem head-on. They are told to increase the number of observations by considering more cases, or decrease the number of variables possibly by collapsing some of them. Alternatively, many qualitative

scholars have argued that a single case should not be mistaken as a single observation. They maintain that since qualitative scholars often measure multiple implications of a theory in a single case, they are in fact obtaining multiple observations from a single case. These extra observations may increase the overall N sufficiently to overcome the problem of indeterminism. Although King et al. cited the strategy with approval, it's not clear to me if measuring extra implications from the same case addresses the statistical indeterminism problem that effectively. As Larry Bartel (2002) remarked, multiple implications of the same theory do not usually belong in the same statistical model – Different explanatory variables may be involved, and different functional forms may be involved. If different implications of a theory are explained by different statistical models, then we have merely replicated the original indeterminism problem to each new model.

The Inferential Logic of Small-N Research

It may be that qualitative scholars have conceded too much in attempting to conform to a statistical logic of inference. For one, the findings from small-N case studies may suggest valid probabilistic causal relationships even when they prove ungeneralizable under a deterministic approach. An interesting example is provided by Sekhon (2002). Using the additional cases furnished by Geddes in her 1990 article, he demonstrated that social revolutions are indeed more likely in countries facing foreign threats. But far more importantly, the investigation of

causality should not be limited to causal effects alone or a statistical paradigm of causal inference. While indeterminism and selection bias are no doubt grievous in the statistical estimation of causal effects, it does not follow that these issues should affect causal process analysis in the same manner.

Indeed, in a classic essay on counterfactual research strategies James Fearon (1991) distinguished between "actual case" statistical regression and "counterfactual" small-N analysis. Although Fearon believed that both approaches follow the same counterfactually defined statistical logic, he maintained that counterfactual causal inference is possible even when there are fewer cases than variables. In "actual case" analysis, the analyst relies on actual observations as substitutes for counterfactual situations. In counterfactual small-N analysis, the analyst makes informed arguments about what *would have* happened. These arguments would be made on the basis of general principles, theories, known regularities and the analyst's intimate knowledge of the cases. The research design is determinate as long as all relevant counterfactual cases are explicitly considered, although as Fearon cautioned "explicit justification of claims about relative effects will require *a proliferation of counterfactual cases*" (Fearon 1991, 178).

Although Fearon seems to regard qualitative counterfactual reasoning as a second-best strategy to be employed when actual case observations are unavailable or unreliable, it is interesting to note that on a philosophical level,

the ability to establish causal relations without appealing to many observations is in fact supposed to be a major advantage of counterfactual causality. Unlike neo-Humean models that rely on patterns of regularity,

> [counterfactual theory] starts with singular events and proposes that causation can be established without an appeal to a set of similar events and general laws regarding them. The possibility of analyzing singular causal events is important for all researchers, but especially for those doing case studies who want to be able to say something about the consequences of Stalin succeeding Lenin as head of the Soviet Union or the impact of the butterfly ballot on the 2000 election (Brady 2002b, 20).

It is clear from these discussions that small-N studies draw their inferential power primarily from causal mechanism analysis. If we acknowledge that a complete understanding of causality involves both effects and mechanisms, then we should not over-emphasize the role of causal effect estimation in the scientific undertaking.[10] And although qualitative scholars should be made aware of the hurdles and pitfalls in assessing causal effects in small-N studies, it would be over-stating the case to insist that case studies

[10] According to the classical Greek conception of science, the mere fitting of events into regular patterns has very little explanatory force. The discovery of descriptive facts (and general laws) is but the first phase of a two-phase process (*ognomy* vs. *ology*). The primary activity of science should be the construction of models of causal mechanisms relating various causal elements. See Stenner, 1998.

cannot produce causal theories. Indeed, it is difficult to imagine how we can investigate causal mechanisms without relying on in-depth small-N studies. From a causal effects standpoint a single case is of little value except in the highly improbable event that it coincides with the statistical average. But from a mechanistic perspective a good case is not necessarily a "typical" case, but a "telling" case which serves to illuminate previously obscure causal connections. The history of science is full of examples in which causal mechanisms discovered in a few or even a single case contributed to important theoretical advances (Stenner 1998, 173-174).

Furthermore, case selection in small-N studies is perhaps best understood from what McKeown calls a "Folk Bayesian" perspective. In the words of Rueschemeyer et al., "as in everyday life we can gain powerful insights from a few encounters because these are assessed against the experience of a life time, so the theoretical framework—when informed by previous thought and research—provides the background against which the picture of the cases studied yields more telling results." (Rueschemeyer, Stephens & Stephens, 1992; p.38). The selection of cases is, therefore, guided by the researcher's prior beliefs regarding the validity of certain theoretical propositions in a given setting. A "least-likely" case, or a case providing what Stephen Van Evera (1997) called a "smoking-gun test" (a case in which the theory generates unique but uncertain predictions), may provide decisive positive evidence for a theory. On the other hand, a "most-

likely" case or a case providing a "hoop test" (a case in which the theory generates close to certain but non-unique predictions) may provide decisive negative evidence against a theory. A "crucial" case is one that provides decisive evidence in both directions. Likewise, comparative strategies such as "most similar" and "least similar" comparisons have a similar logic.

This is certainly not to say that small-N scholars should not be wary of selection bias in their research design. But we should be equally wary of too dogmatic an application of statistical principles to small-N research. It is not especially relevant to argue in terms of OLS curve fitting when the investigator is interested primarily in untangling causal processes. From a Bayesian perspective there is no knowledge without foreknowledge. And a comparative researcher who selects on the dependent variable (X's) for actual observations may still be able to construct counterfactual cases (not X's) based on prior knowledge and causal process analysis. It's not clear if and to what extent process tracing can mitigate the selection bias problem. But we probably should not be too quick to conclude that process tracing is no help at all.

Finally, a few words should be said about the use of Millian methods in small-N studies. Critiques of Mill's methods have usually focussed on the three very strong conditions necessary for their application. These are: 1) The absence of equifinality—In other words the causal relations must be deterministic regularities involving necessary and/or sufficient Conditions: 2) The

identification of all relevant variables; 3) The presence of cases representing the full range of possible causal paths. But as Bennett and George (1997) argued in their excellent exposition on the topic, causal process analysis in small-N studies is in fact designed to address these limitations head-on. By teasing out multiple causal pathways and examining the interaction of causal variables in depth, process tracing addresses the equifinality issue directly. By allowing for the "inductive identification of variables as well as their deductive specification," it reduces the likelihood of omitted variables. Lastly, process tracing helps researchers construct informed counterfactual arguments when all possible types of cases are not available. Bennett and George emphasized that case study methods should not be conflated with the Millian methods, as the latter generally only serve to structure process tracing in the former.

Enhancing Inferential Power in Small-N Research

Advice to qualitative scholars on how to strengthen causal arguments in small-N studies must take into account the epistemological basis of qualitative methods. For making causal inference small-N researchers often rely on causal process analysis, frequently structured by the Millian methods. Insights from quantitative methods can be enlightening, but a naïve transplantation of statistical principles to small-N studies can be misleading and counterproductive.

Small-N researchers should be aware of the strengths as well as the weaknesses of their methodology and be conscious if not necessarily explicit about the model of causal inference to which they appeal. In-depth small-N analysis lends itself to process tracing and the discovery of causal mechanisms, but is less effective for obtaining reliable measurements of causal effects. Nor is it suited to a neo-Humean logic of inference emphasizing patterns of regularity. To leverage the strengths of the small-N approach researchers should emphasize the investigation of causal mechanisms while acknowledging that estimates of causal effects obtained from a few cases are rough and uncertain. Mill's methods can provide a useful framework for analysis but their limitations must be corrected by careful process tracing.

Some useful guidelines for process tracing as a tool of causal inference can be found in the works of the historian Clayton Roberts (1996). Like Stuart Glennan, Roberts emphasized that each step in the causal process must be subsumed within an appropriate covering causal law. Within this framework the more micro the causal mechanism, the more precisely the causal laws covering it can be defined and the easier they can be verified. Ideal-typical process tracing therefore involves the minute tracing of the explanatory narrative to the point where the explanatory mechanisms are "microscopic" and the covering laws are correspondingly more clearly defined and readily verifiable. For the empirical investigator, the important lesson seems to be that one should strive to construct

causal arguments on a foundation of robust causal generalizations, achieving greater certainty by deepening the level of analysis whenever necessary. Although the verification of every causal generalization in an inductive study cannot possibly be expected, one should at least be explicit about the untested generalizations on which a causal argument hinges. Process tracing is subject to the same standards of falsifiability as any other scientific procedure.

However, small-N researchers need not appeal exclusively to a mechanistic model of causality. Counterfactual arguments can provide clarity and extra persuasive power to causal process analysis. As James Fearon suggested, small-N researchers "could strengthen (or simply specify) their causal arguments by being *explicit* about the counterfactual scenarios needed to support their hypotheses.… If, for whatever reason, one is reluctant to add actual cases, then it is essential to make explicit what might have happened if a possible cause had varied" (Fearon 1991, 194). Convincing causal arguments must be built on a solid foundation of compelling theories, reliable patterns of regularity, and careful identification and observation of causal mechanisms. In addition, both qualitative as well as quantitative analysts should pay more careful attention to the *contenability* of counterfactual conditions. That is to say, one should always be sensitive to the possibility that assigning a counterfactual value to one variable may render the overall case configuration implausible.

Lastly, more encouragement should be given to the appropriate incorporation of large-N results and methods into small-N studies and the synthesis of qualitative and quantitative research in general. Statistical methods can be employed by small-N scholars to great advantage in the identification of research problems, and no doubt more reliable estimates of causal effects can only facilitate the explication of causal mechanisms. In certain cases, process tracing may reveal critical links in the chain of causal explanations and the cogency of the overall argument can be significantly strengthened if these putative connections are then borne out with large-N data. We don't need to subscribe to a statistical worldview, but we should appreciate that the sociopolitical causal machine may be held together by statistical glue.

Conclusion

Regardless of the methodology and model of causality chosen, a central fact confronting all empirical political scientists is the scope and impact of human agency in political phenomena. One needs not subscribe to a predominantly "cloud-like" conceptualization of the social world to acknowledge that in the realm of politics, most regularities are soft and many have short half-lives. As Almond and Genco (1977) pointed out, they are soft because they are embedded in a complex web of human impulses and intentions. They decay quickly "because of the memory, creative searching, and learning that underlie them. Indeed, social science itself may contribute to this decay, since learning

increasingly includes not only learning from experience, but from scientific research itself" (Almond and Genco 1977, 494). If dictators also subscribe to the *Journal of Democracy*, how confident can we be that past patterns of democratization will hold in the future? The scenario is not as facetious as it may appear. The Communist leadership of China, for example, had shelved plans for the expansion of direct local elections, owing in no small part to a growing recognition of the role of such elections in the disintegration of authoritarian regimes elsewhere in East Asia. In a world populated by social actors who constantly react and adapt to observed patterns, the very act of observation may fundamentally alter the phenomenon of interest. As Glenn Shafer (2000) cautioned, we can only expect very nuanced answers to counterfactual questions, and it is not at all clear how or to what extent can observed causal relations—statistical or otherwise—be extrapolated.

This is a basic dilemma from which there is no complete escape. But we can probably deal with it more effectively if we avail ourselves of the diverse tools in our repertoire. A "top-down" approach emphasizing general laws and explanatory parsimony is not enough by itself. One would also need to call upon a "bottom-up" approach emphasizing mechanisms to study singular events, human learning and innovation, and the workings of Popperian "plastic

control."[11] In the words of Brady, "there is no need to enshrine one approach over the other. Both have their uses." And they may in fact be indispensable to each other.

[11] Karl Popper uses the concept of "plastic control" to incorporate the role of free will into human control of behavior and other aspects of the physical world. Cited in (Almond & Genco 1977).

References

Achen, Christopher H., and Duncan Snidal. 1989. "Rational Deterrence Theory and Comparative Case Studies." *World Politics* 41: 144-69.

Adcock, Robert; Collier, David. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research", 95(3) *APSR,* 529-546.

Almond, Gabriel A, and Stephen J. Genco. 1977. "Clouds, Clocks and the Study of Politics." *World Politics* 29: 489-522.

Bartels, Larry M. 2002. "Some Unfulfilled Promises of Statistical Imperialism", in Brady,

Henry E., and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Berkeley, CA: Rowman & Littlefield Publishers, Inc. and Berkeley Public Policy Press.

Bennett, Andrew and Alexander George. 1997. "Process Tracing in Case Study Research," Paper presented at the MacArthur Foundation Workshop on Case Study Methods, Belfer Center for Science and International Affairs (BCSIA), Harvard University, October 17-19, 1997.

Bennett, Andrew. 2002. "Where the Model Frequently Meets the Road: Combining Statistical, Formal, and Case Study Methods". Paper presented at the Annual Meetings of the American Political Science Association, Boston, MA 2002.

Brady, Henry E. 2002a. "Doing Good and Doing Better: How Far Does a Quantitative Template Get Us?" in Brady, Henry E., and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press, 2002.

Brady, Henry E. 2002b. "Models of Causal Inference: Going Beyond the Neyman-Rubin-Holland Theory", paper presented at the Annual Meetings of the Political Methodology Group, University of Washington, Seattle, Washington, July 2002.

Brady, Henry E.; Collier, David; Seawright, Jason. 2002. "Refocusing the Discussion of Methodology", in Brady, Henry E., and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Berkeley, CA: Rowman & Littlefield Publishers and Berkeley Public Policy Press.

Collier, David. 1991. "The Comparative Method: Two Decades of Change, "in *Comparative Political Dynamics*, edited by Dankwart A. Rustow and Kenneth Paul Erickson. New York: Harper Collins, pp.7-31.

Collier, David; Brady, Henry; Seawright, Jason. 2002. "Eliminating Rival Explanations: Evaluating Sources of Leverage", in Brady, Henry E., and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press.

Collier, David, and James Mahoney. 1996. "Insights and Pitfalls: Selection Bias in Qualitative Research", *World Politics* 49, no. 1: 56-92.

Collier, David; Seawright, Jason; Munck, Gerardo. 2002. "The Quest for Standards: King, Keohane, and Verba's *Designing Social Inquiry*", in Brady, Henry E., and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press.

Dawid, A. P. 2000. "Causal Inference Without Counterfactuals," *Journal of the American Statistical Association*, Vol. 95, No. 450: 407-448.

Fearon, James D. 1991. "Counterfactuals and Hypothesis Testing in Political Science," *World Politics*, Vol. 43, No. 2: 169-195.

Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis* 2: 131-50.

Gerring, John. 2001. *Social Science Methodology*. Cambridge: Cambridge University Press.

Glennan, Stuart. 2000. "Rethinking Mechanistic Explanation," working paper, http://hypatia.ss.uci.edu/lps/psa2k/glennan.pdf.

Hitchcock, Christopher. 2002. "Probabilistic Causation", in *Stanford Encyclopedia of Philosophy* (Fall Edition), Edward N. Zalta (ed.) http://plato.stanford.edu/archives/fall2002/entries/causation-probabilistic/.

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

Lieberman, Evan. 2002. "Seeing Both the Forest and the Trees: Nested Analysis in Cross-National Research." Paper presented at the Annual Meetings of the American Political Science Association, Boston, 2002.

Lieberson, Stanley. 1992. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases," in Charles Ragin and Howard Becker (eds.), *What is a Case? Exploring the Foundations of Social Inquiry*. Cambridge: Cambridge University Press, pp. 105-118.

Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65, no. 3: 682-93.

Mahoney, James. 1999a. "Nominal, Ordinal, and Narrative Appraisal in Macrocausal Analysis." *American Journal of Sociology* 104, no. 4: 1154-96.

Mahoney, James. 1999b. "Strategies of Causal Inference in Small-n Analysis," *Sociological Methods and Research*, 1999.

McKeown, Timothy. 1999. "Case Studies and the Statistical Worldview." *International Organization* 53, no. 1: 161-90.

Most, Benjamin A.; Starr, Harvey. 1989. *Inquiry, Logic and International Politics*. Charleston: University of South Carolina Press.

Munck, Gerardo L. 2002. "Tools for Qualitative Research", in Brady, Henry E., and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press.

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

Ragin, Charles. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

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

Rueschemeyer, Dietrich. 2002. "Can One or a Few Cases Yield Theoretical Gains?" in *Comparative-Historical Analysis in the Social Sciences*, edited by James Mahoney and Dietrich Rueschemeyer, 305-36. New York: Cambridge University Press, 2002.

Rueschemeyer, Dietrich; Stephens, Evelyne Huber; Stephens, John D. 1992. *Capitalist Development and Democracy*. Chicago: University of Chicago Press.

Sekhon, Jasjeet Singh. 2002. "Revisiting Case Selection: Conditional Probability and John Stuart Mill's Methods of Inductive Inference". Working paper. http://jsekhon.fas. harvard.edu/papers/mill.pdf

Shafer, Glenn. 2001. "Nature's Possibilities and Expectations," in *Probability Theory: Philosophy, Recent History and Relations to Science*, Vincent F. Hendriks, Stig A. Pedersen, Klaus F. Jorgensen (eds.), pp. 147-166. Kluwer: Synthese Library, Vol. 297.

Sil, Rudra, and Katzenstein, Peter. 2005. "What Is Analytic Eclecticism and Why Do We Need It?" Paper presented at the Annual Meeting of the American Political Science Association, 2005.

Skocpol, Theda, and Margaret Somers. 1980. "The Uses of Comparative History in Macrosocial Inquiry." *Comparative Studies in Society and History* 22, no. 2: 174-97.

Sprites, Peter; Glymour, Clark; Scheines, Richard. 2000. *Causation, Prediction and Search*, 2nd edition. Cambridge, MA: M.I.T. Press.

Stenner, Paul. 1998. "Heidegger and the Subject Questioning Concerning Psychology", *Theory & Psychology*, Vol. 8, # 1.

Taylor, Charles. 1977. "Interpretation and the Sciences of Man," in Fred Dallmayr and Thomas A. McCarthy (eds.), *Understanding and Social Inquiry*, (South Bend: University of Notre Dame Press), pp.101-131.

Verkuilen, Jay. 2002. Manuscript for Ph.D. dissertation.

Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. Ithaca: Cornell University Press.