Spring 2005, Vol. 3, No. 1

Qualitative Methods

Newsletter of the
American Political Science Association
Organized Section on Qualitative Methods

Contents

APSA-QM Section Officers

Letter from the Editor

John Gerring
Boston University
jgerring@bu.edu

In this issue, I implored authors to cast aside the usual norms of comity and good taste and, instead, to engage each other directly and unsparingly, leaving no reputation intact. Goading them on in this unscrupulous fashion I fully anticipated making my reputation in the broader world as an editor provocateur, with attractive options in the commercial world of publishing and cable television (*Esquire*? Crossfire? The O'Reilly Factor?). I had my sights set.

Unfortunately, the authors insisted upon respecting each others' opinions, even as they thrust and parried. Readers of this issue will find that the two symposia resemble not mud-wrestling but rather equestrian battles, with all due norms of civility observed. Perhaps, in the end, it is more edifying, as well as more trenchant. I do not wish to de-fang the debates...

The first symposium is a wide-ranging (though by no means comprehensive) collection of views on the qualitative/quantitative distinction. Gerry Munck begins with a strong critique of qualitative methods, a field that "rests on a faulty methodological foundation." Ten fallacies inhibit the development of a consensus about what constitutes good methodological practice among qualitative researchers. Andy Bennett takes issue with each of Munck's criticisms, defending the progress that has been made within what might be called (here I resume my role of methodological provocateur) the "qualitative template." Ken Benoit pursues a line of argument that is, depending upon the reader's perspective, an extension of Munck's. Where it is possible to count things, Benoit argues, we ought to do so, for there are many methodological benefits to quantitative research. More important, there is no significant distinction between these two (supposed) forms of knowledge; words are incipient numbers. When we have several similar things we can—in addition to calling them by names—also count them. This offers distinct advantages, in addition to parsimony. Bernhard Kittel strikes out on many fronts, in an attempt to summarize various differences between American and European methodological perspectives. His piece offers a counterpoint to Benoit's, since Kittel—like Benoit, known mostly for quantitative work—is much more critical of the quantitative template (as articulated by King, Keohane, and Verba). Finally, Jim Mahoney defends, or at least re-articulates, several aspects of Kittel's criticism that are focused explicitly on comparative-historical analysis. Thus, readers will find a wide range of views, emanating from scholars who are often critical of the research traditions with which they are most familiar and with which their own work is most often associated. (There is something noble in an equestrian battle in which the knights carry the other monarch's standard. But nothing that Bill O'Reilly is likely to appreciate.)

The second symposium focuses on a smaller target, the debate over necessary conditions, which has nonetheless become quite a heated issue among qualitative and quantitative scholars. Gary Goertz, who is among the leading spokespersons for this view of causal relations, sets forth to counter the prevailing view that necessary conditions must be interpreted as "deterministic." David Waldner responds that if necessary does not mean deterministically necessary then it introduces a fundamental—and unnecessary—ambiguity into methodological discussion. Jas Sekhon, finally, is skeptical of the possibility of arriving at a probabilistic analysis of necessary conditions unless there is also an explicit attempt to incorporate uncertainty (aka error). Causal analysis in a probabilistic world assumes sample distributions and these distributions must be modeled, or at least discussed.

On your marks, get set...

Symposium I: The Quantitative/Qualitative Distinction

Ten Fallacies About Qualitative Research

Gerardo L. Munck[1]
University of Southern California
munck@usc.edu

"Words *alone* beat numbers alone.
Words *with* numbers beat words alone.
And numbers make sense, or much greater sense,
within verbal theory" (Sartori 1976: 319).

Qualitative research, defined here in contrast to quantitative research as consisting of verbal as opposed to numerical statements or, more simply, of words as opposed to numbers, is an inextricable, necessary component of the social sciences. Moreover, for a variety of reasons, the bulk of existing knowledge in the social sciences has been generated through qualitative research and this form of research probably will continue to be the most commonly used path to knowledge. Yet a great part of the potential of qualitative research is not realized because the methodological foundation of this research is shaky.

To substantiate the claim that much qualitative research rests on a faulty methodological foundation, this essay discusses various fallacies *about* qualitative research found in the literature on qualitative methodology that researchers turn to for guidance. That is, the focus is not on fallacies *in* qualitative research, that is, instances where researchers depart from sound methodological principles. Indeed, the source of problems is deeper and can be traced back to some widely held myths about the methodology of qualitative research.

The essay is organized under two headings. It focuses initially on descriptive theorizing, which is concerned with forming and measuring concepts. Subsequently it turns to causal theorizing, which is concerned with hypothesizing and testing cause-effect relationships among concepts. Ten fallacies are identified and discussed. But some of the fallacies are quite general and crop up, if in a different form, at various points in the research process. Thus, strictly speaking, rather than spotlight ten fallacies about qualitative research, this essay addresses a range of fallacies associated with ten methodological issues, each of which is fundamental to the conduct of qualitative research.

The discussion is not exhaustive and it should not be read as indicating either that these fallacies are exclusive to the methodological literature about qualitative research or that there is not much that is useful and important in this literature. The intent of this essay is not to offer a comprehensive assessment of current work on methods used in qualitative research, let alone an assessment of the relative strengths and weaknesses of methods used in qualitative and quantitative research. Rather, the point is to take a hard look at the advice on central methodological issues routinely given to qualitative researchers and to draw attention to erroneous views about the methodological underpinnings of qualitative research. The promotion of good qualitative research obviously requires positive guidance regarding how such research should be conducted; for reasons of space, these lessons can only be hinted at in this essay. Nonetheless, identifying and unlearning mistakes is part and parcel of the difficult but fascinating pursuit of knowledge.

Descriptive Theorizing: Concepts and Measures

1. The Ladder of Abstraction, Conceptual Stretching, and the Fallacy of a Priori Domain Restrictions. In a classic text for qualitative methodologists, Giovanni Sartori wrote about a ladder of abstraction or more correctly, as has been pointed out, a ladder of generality that is the foundation for what he later codified as rule Nº 7 of concept formation. This rule states that "the connotation [i.e. the concept's attributes or content] and the denotation [i.e. the concept's empirical referents] of a concept are inversely related." (1970: 1040-46, 1984: 44). This rule has many implications, the most important of which is that inherent in the logic of concept formation is a trade-off between large-N and small-N research. Large-N researchers run the risk of engaging in conceptual stretching—understood as the application of a concept to a case that does not possess the attributes used to define the concept—unless they employ thin concepts, that is, concepts that have few attributes. In turn, analysts who want to use thick concepts—that is, those with many attributes—must avoid conceptual stretching by conducting small-N research.

Qualitative methodologists have frequently repeated Sartori's rule and taken its implication for small-N research to heart. Yet this justification for small-N analysis rests on the fallacy of a priori domain restrictions. First, statements about the distinctiveness of a set of cases or, at the extreme, the uniqueness of any case should not be tackled prior to coding, at a purely conceptual level, and should not be invoked as a basis to shortcut empirical research. Rather, such statements are only valid as conclusions about a case or set of cases that emerge from an empirical analysis. Second, the domain of applicability of a concept is not cases that possess a certain attribute, as is suggested by cases that emerge from an empirical analysis a *per genus et diffentiam* mode of analysis, let alone cases that are similar, as is proposed by advocates of a restrictive understanding of comparability. The domain of a concept extends to all units that have the potential to possess a certain attribute; in other words, it extends to negative cases. Thus, to give some examples, one should not exclude from one's universe of cases instances of authoritarianism because one is interested in democracy, of revolution because one is concerned with stable polities, or of failed states because one seeks to address successful ones. Indeed, the basic point about systematic description is to devise instruments that allow analysts to compare cases by identifying both similarities and differences in terms of a common metric.

2. Differences of Kind vs. of Degree, and the False-Dichotomy Fallacy. Qualitative methodologists have argued that, when it comes to measuring concepts, a key choice is whether to cast distinctions as differences of kind or differences of degree. Moreover, differences of kind are portrayed as qualitative differences and are seen as more fundamental than, and in some sense overriding, quantitative differences of degree. Finally, because differences of kind and degree are seen as mutually exclusive options, researchers are seen as facing a trade-off that offers a justification for qualitative measures. At least for certain research purposes, the standard hierarchy of levels of measurement is overturned and qualitative differences of kind are held to be superior to quantitative differences of degree; to use terms common to this literature, dichotomous measures are seen as superior to continuous measures.

The problem is that the posited trade-off used to opt for measures that highlight differences of kind over differences of degree is based on a false-dichotomy fallacy. All measurement involves first and foremost classification, based on dichotomous distinctions between cases that are relatively similar to each other and relatively different from other cases in terms of some category. Indeed, the most basic decision in measurement, which underlies even the most powerful and sophisticated measures, is the drawing of a boundary, a line on a continuum, that establishes an equivalence/difference relationship. But all measurement can also be equated to quantification, in that it can be seen as consisting of assigning numbers to objects according to rules. And there is nothing that prevents insights about differences in kind from being translated into numbers. Thus, the proper distinction to be made among scales concerns the mathematical properties of the relationships among the numbers used in each scale. And, assuming that all measures are of equal validity, it follows that higher-level measures are always preferable in that they offer more information than lower-level scales.

3. Aggregate Measures and the Fallacy of Conflation. Qualitative methodologists distinguish two key steps in the process of measurement: the identification of attributes used to define a concept and the assignment of scores to cases on these attributes. Yet even though the concepts discussed by qualitative researchers usually have more than one attribute, and even though the measures assigned to these attributes are frequently aggregated into a single summary score, these methodologists do not routinely distinguish between disaggregate and aggregate measures and tend to conflate these two distinct orders of measures. Examples of this failing are found in many methodological discussions regarding the validity of measures of democracy and political regimes.

The tendency to conflate disaggregate and aggregate measures has three key implications. First, the need to explicitly address and justify the choice of aggregation rule, a key step in the process of measurement, is not recognized. Second, no distinction is drawn between the impact on aggregate scores due, on the one hand, to choices regarding the scales used to measure each conceptual attribute and, on the other hand, to choices regarding the aggregation rule. Third, the need to avoid terminological confusion by distinguishing the labels used to refer to values of the aggregate and disaggregate scales is not highlighted. In sum, the literature on qualitative methodology fails to propose procedures that would distinguish among different orders of measures and to offer a basis for generating readily interpretable aggregate measures.

4. Objectivism vs. Subjectivism, and the Fallacy of Appeals to Authority. Qualitative methodologists routinely reject positivism, drawing attention to the theory-laden nature of all observations. However, the correct assertion that there is no objective reality, that the world is always seen from a particular position, and that all data must be interpreted is frequently used to set up a contrast between objectivism and subjectivism, and to justify subjectivism. In this view, since there is nothing objective, descriptions cannot escape "he said, she said" kinds of exchanges and cannot adjudicate among rival views of the world. Moreover, because scientific description is not seen as possible, debates get resolved ultimately on the basis of appeals to the authority of the person emitting an opinion, the invocation of science being but one way—a dishonest one—to mobilize ideological power behind certain opinions.

This anti-scientific view is misguided. Indeed, it is a fallacy to appeal to authority, and positing an equivalence between such appeals and science fails to capture certain distinctive features of the process of scientific inquiry. Scientific description takes theory as its point of departure. But social scientists avoid subjectivism in two important ways. First, they do so by ensuring that their methods are replicable. This entails, at its core, the generation of measures on the basis of clearly specified rules and observables. In this way conclusions are not presented under the guise of any authority but rather are open to a process of independent inspection, whereby other analysts are allowed and even invited to consider whether they arrive at the same conclusions. Second, they avoid subjectivism by exposing parts of their theoretical assumptions—ideally ever increasing parts—to tests by framing issues as empirical questions that admit right or wrong answers. In a nutshell, there is such a thing as a social science.

Causal Theorizing: Causal Models and Tests

5. Testing with an N=1, and the Fallacy of Hasty Generalization. Qualitative methodologists frequently defend the potential contributions to causal assessment of a case study, that is, a study with an N=1. The considerable value of knowing a case notwithstanding, the claim that causal theories can be tested with an N=1 is an example of the fallacy of hasty generalization. Of course, it is possible to have many observations even with an N=1 study. Nonetheless, inasmuch as the purpose of a study is to say something about processes that operate at the level of the unit that defines the N (a country, for example), a case study has limited value as a means to test hypotheses. To draw conclusions from an N=1 study at the very least three assumptions must be made: (i) causation operates deterministically and there is no plausible probabilistic alternative hypothesis; (ii) a complete theory that includes all the variables needed to explain all the variation in the outcome of interest is available; and (iii) all the variables are measured without error. And each of these three conditions rests on extremely implausible assumptions that can be relaxed only by increasing the N.

6. Testing Causal Mechanisms, and the Fallacy of Distraction. Qualitative methodologists are keen to draw attention to a key feature of case studies: their sensitivity to process. This is no doubt true and this characteristic makes case studies particularly suited for gathering data relevant to theories that posit causal mechanisms, that is, that are framed in terms of actors and actions. But these methodologists take a false step when they use this feature to respond to critiques of N=1 or small-N studies as means to test causal theories and, going even further, to advance the claim that it allows qualitative researchers to get around degrees of freedom problems that cannot be solved through large-N analysis.

Arguments framed in terms of causal mechanisms have no distinct epistemological status and advantage, and must be treated simply as one type of hypothesis to be tested against other plausible hypotheses. Thus, this defense of qualitative research as a means of causal assessment hinges on the fallacy of distraction, the illegitimate changing of the topic of discussion to deflect attention from the weakness of one's position. It is, no doubt, important to test theories with valid measures and case studies are an indispensable means, indeed probably *the* most suitable, of gathering information relevant to causal mechanisms. But this virtue does not get around the need to base causal assessment on reasonable assumptions and hence, as suggested in the previous point, to increase the N.

7. Case-oriented vs. Variable-Oriented Approaches, and the False-Dilemma Fallacy. Qualitative methodologists have claimed that qualitative researchers use a distinctive case-oriented approach to causation that, in contrast to a variable-oriented approach, considers potential interactions among causal factors and, more broadly, has the virtue of being attuned to the integrity of cases, that is, to the way variables generate effects operating as parts of an indivisible whole. These are important, desirable features. Nonetheless, this contrast between case- and variable-oriented approaches, as other attempts to identify distinctive features and strengths of qualitative research, is fallacious, because it hinges on a false dilemma.

All causal theorizing, inasmuch as it is analytic, breaks down a problem into constituent parts. These parts are called, by convention, variables. And cases are but units that are characterized by a bundle of specific values assigned to all the variables in a causal model. Thus, researchers are not forced to choose between studying cases or variables. Indeed, there is no methodological foundation to the choice between case- vs. variable-oriented approaches or, as the choice is sometimes framed, between holistic vs. analytical approaches. The decision to consider the relationship among causal variables, that is, to emphasize the way in which the parts fit together to constitute the whole, as opposed to seeing these variables as operating in an isolated fashion, is merely a choice regarding the causal model.

8. Context and the Fallacy of Ad Hoc Explanations. Qualitative methodologists suggest that case studies bring to light contextual factors that may well vary from case to case and have an impact on outcomes, frequently through their interactions with the factors highlighted in causal theories. Sensitivity to context is, without doubt, a central feature and virtue of case studies. But, from a methodological perspective, there is a generally unrecognized downside to the way in which recourse to context introduces too much flexibility into the research process. Indeed, the invitation to draw on explanatory factors from outside the researcher's causal model and the failure to insist on the need to formalize such contextual factors in one's model and to retest the model, amounts to a severe underappreciation of the danger of the fallacy of ad hoc explanations, one of the most common fallacies in qualitative research. In effect, the trumpeting of the benefits of considering context is usually not balanced with an equally clear recognition of the way in which such inductive thinking voids the value of using such case studies as a means of theory testing.

9. Causal Model Specification, Estimation Methods, and the Fallacy of Conflation. Qualitative methodologists discuss a range of issues researchers must address, and a range of tools that researchers can use, when they seek to test theories with case studies. Yet these discussions regularly conflate distinct options faced by researchers, most obviously with regard to the specification of a causal model. For example, the common assertion that necessary conditions are an example of causes that operate deterministically conflates two distinct aspects of a causal model: (i) the predicted pattern of the data (e.g., linear or non-linear patterns, necessary conditions being an example of a non-linear pattern), and (ii) the tolerance for deviations from the predicted data pattern (e.g., deterministic or probabilistic standards). The fallacy of conflation also crops up in discussions about methods for estimating a causal model's parameters. For example, while it is quite standard for qualitative methodologists to discuss causal models that entail non-linear effects and interactions among causal factors, with only few exceptions the range of methods for estimating a causal model's predictive power never gets beyond a vague notion of association that actually smuggles in, as a default option, assumptions of linearity and additivity. Whether explicitly or implicitly, key choices faced by researchers in specifying and estimating their causal model are not identified and distinguished with clarity.

10. Cumulation and the Questionable-Cause Fallacy. Finally, qualitative methodologists regularly suggest that qualitative research can lead to conclusions about causal relationships as well as to cumulative knowledge about such relationships. These claims are made in the absence of summary measures of causal effect, estimates of the uncertainty and likely bias of results, procedures to correct for selection bias, tests of robustness, and procedures to carry out a meta-analysis. Thus, it is hard to see how such claims can avoid the questionable-cause fallacy, that is, the mere assertion that a certain factor is a cause when it has not actually been shown to be a cause. In effect, such claims largely gloss over the difference between informed opinion and the results of scientific inquiry, and no doubt lead to the acceptance of erroneous hypotheses as well as the rejection of correct hypotheses.

Conclusion

This essay has argued that the current literature on qualitative methodology contains much fallacious advice and, moreover, that these fallacies pertain to some of the main methodological issues that need to be addressed in conducting qualitative research. This does not mean that there is not much that is sound and useful in this literature. Moreover, the views presented here could very well be contested. But, in raising these issues, this essay will hopefully have succeeded in justifying the need for a thorough debate over the conventional methodological wisdom about qualitative research.

Arguing that current discussions about qualitative research contain fallacious methodological advice might fuel the already considerable skepticism among many qualitative researchers about the value of learning about methods. Any investment in developing methodological skills constitutes a diversion from energies dedicated to substantive research. Disputes about methodology only add weight to the view that researchers should just rely on their intuition, which has served them well for a long time, and simply sidestep complex debates with unclear payoffs. This is true. But the conclusion that methodology should be disregarded does not follow from these premises. It is critical that qualitative researchers acknowledge that substantive arguments are only as good as the methodological foundation upon which they rest and understand the grounds for making knowledge claims.

Notes

[1] I would like to thank Michael Coppedge, John Gerring, Angela Hawken, Staffan Lindberg, Sebastián Mazzuca, Aníbal Pérez-Liñán, Andreas Schedler, Richard Snyder, Saika Uno, and Jay Verkuilen for comments, but in no way want to suggest that they all agreed with all of the points in this essay.

References

Sartori, Giovanni. 1970. "Concept Misinformation in Comparative Politics," *American Political Science Review* Vol. 64, No 4 (December): 1033-53.

Sartori, Giovanni. 1976. *Parties and Party Systems. A Framework for Analysis.* (Cambridge: Cambridge University Press).

Sartori, Giovanni. 1984. "Guidelines for Concept Analysis," pp. 15-71, in Giovanni Sartori (ed.), *Social Science Concepts. A Systematic Analysis.* (Beverly Hills, Cal.: Sage Publications).

The Fallacy of Fallacies

Andrew Bennett
Georgetown University
BennettA@Georgetown.edu

Gerardo Munck has done the field a service by raising fundamental critiques of a broad range of topics in the burgeoning literature on qualitative research methods. In view of the remarkable outpouring of books and articles on these methods in the last decade, it has no doubt been difficult for practitioners of qualitative methods, and even for methodologists of all stripes, to keep up with and make sense of the large number of sometimes competing claims raised in this fast-moving literature. Munck's critiques are timely in this regard, and an ongoing discourse in which qualitative methodologists are among their own toughest critics is a welcome sign of the heightened state of development that qualitative methods have achieved.

Any single reading of this new literature is necessarily open to contention. Without purporting to have the last word on the subject, I offer my own reading regarding the issues Munck raises, emphasizing the points on which we differ in order to be as informative as possible. On the whole, while I agree that there is a danger that the new literature can and no doubt will be misused to justify unwarranted inferences and research practices, I argue that on some issues much of the new literature is attuned to and warns against precisely the kinds of potential fallacies that concern Munck, while on other issues the new literature is aware of the dangers Munck raises but is concerned as well about competing concerns or trade-offs. Munck is perhaps trying to be (unnecessarily) polite in not citing specific methods texts that fall prey to the "fallacies" he cites, but this raises confusion as to which writings he has in mind. I cite specific writings (some of them new or forthcoming and thus potentially not yet available to Munck) that share and address his concerns on some points and take a different view on others. I address each of Munck's points below using Munck's headings to facilitate comparison by the reader, though it should be clear that I would use different section headings if I were writing de novo (using, for example, the term "danger" or "inferential risk" instead of "fallacy," as I believe the extant literature often recognizes the problems that Munck raises).

1. The Ladder of Abstraction, Conceptual Stretching, and the Fallacy of a Priori Domain Restrictions. Munck is right to note that there is a danger of arbitrarily or prematurely restricting the domain of a theory to a set of cases without giving serious thought to the population of "negative cases," or potential cases that should be within the domain of the theory and are relevant to testing it and specifying its scope conditions. This has been a central concern in the new literature, however. Ragin (2000) devotes great attention to problematizing the population to which theories apply and notes that the specification of such populations usually emerges only after much research, not as a prior step. Indeed, Ragin emphasizes that a major distinction between qualitative and statistical methods is that the former do not start out with a fixed population about which inferences are to be drawn. Collier and Mahoney (1996) focus on the close connection between the cases selected for study and the generalizations to specified populations that are defensible in view of the cases selected and the empirical findings. George and Bennett (2005) note in the context of the subject of typological theorizing that qualitative research often involves iteration among evidence from individual cases, the theories used to explain these cases, and the cases selected for study or the population to which the theory should apply, and they offer standards to avoid "curve-fitting" in this iterative process (most notably the injunction to look for new observable implications, both within cases and across cases, of any changes to the theories being tested). Finally and most specifically, Mahoney and Goertz (2004) have written on the challenge of identifying negative cases and suggested a (contestable) rule for specifying the population of negative cases. While there are many differences among these authors' writings on this subject, none of them suggest that the domain of theories should be specified prior to considerable research—indeed, the goal of "testing" theories in case studies is as much to specify the conditions under which they apply and identify or verify the processes through which the outcome arose as it is to reach assessments on the "general" validity of theories for broad populations. Put another way, qualitative researchers are often as interested in the "causes of effects," and the contexts in which these causes do and do not operate, as they are in the "effects of causes" across broad populations; qualitative methods have advantages for getting at the first of these aspects of causality, and quantitative methods have advantages for getting at the second.

2. Differences of Kind vs. of Degree and the False-Dichotomy Fallacy. In my view qualitative methodologists don't cast our measurement choices as having to be limited to differences in kind or differences in degree. The title of Jim Mahoney's 1999 article captures the wide options available to qualitative researchers: "Nominal, Ordinal, and Narrative Appraisal." Munck is certainly right that a key issue is that of drawing boundaries or partitions based on differences of degree or kind that specify what kind or level of a variable is considered to be comparable across cases or subject to unit homogeneity. Identifying such partitions is often an important research goal. Munck is right that higher-level measures can offer more information than lower-level scales, and qualitative researchers have the option of using such measures. As Ragin (2000) notes, however, higher-level measures can seem to offer more information than they actually do when values at the very high or very low end of a scale are not meaningfully related to the question being investigated—beyond a certain point, being even more pregnant, drunk, or mortally wounded does not matter much for many intents and purposes. This builds on Paul Lazarsfeld's insight more than fifty years ago that sometimes the extreme ends of a scale can be collapsed down into categorical measures without any meaningful loss of information if the variance at the extreme ends of the independent variable is not relevant to the subject being investigated. Translating nominal variables into quantitative measures in a way that adds information is also problematic (they can be converted into dummy variables, of course, but this adds no information). In short, while quantifying some variables can indeed add information, it may not add information to quantify other variables, and with some variables quantifying variance without careful consideration of whether all of the variance is meaningful can reduce the internal validity of the measures used.

3. Aggregate Measures and the Fallacy of Conflation. Munck argues that "methodologists do not routinely distinguish between disaggregate and aggregate measures and tend to conflate these two," and he suggests that this failing is common in measures of democracy and political regimes. Munck has done outstanding work on measures in this literature that he is perhaps too modest to cite (Munck and Verkuilen, 2002), and I do not doubt that there are examples of this problem within this area. I would only point out that in the more general literature on qualitative methods, scholars have paid a great deal of attention to this and many other issues of conceptualization and measurement, including Adcock and Collier (2001), and Goertz (forthcoming).

4. Objectivism vs. Subjectivism, and the Fallacy of Appeals to Authority. Certainly some qualitative methodologists reject "positivism," but this term has been subjected to so many meanings as to be meaningless—indeed, if by "positivism" Munck means "logical positivism," then no one (including Munck) defends logical positivism. It is impossible to do justice to the wide range of epistemological views among qualitative methodologists in this brief forum, so I limit myself to noting that none of the methodologists I have cited thus far would argue that there is no such thing as social science, and all would accept Munck's two standards of seeking replicable methods and subjecting theories to empirical testing. Indeed, many of those cited herein would add more ambitious epistemological goals for qualitative research and have engaged in extended discussions of fundamental issues in the philosophy of science (see, for example, George and Bennett 2005).

5. Testing with an N=1, and the Fallacy of Hasty Generalization. I will happily plead guilty to being one of those who has defended the potential for theory-testing even in single-case studies. This should not be misconstrued as arguing that single case studies are always, usually, or often the best means of testing theories. Yet there are contexts and ways in which single-case theory testing can be valuable. If we assume the world to be probabilistic and if we construe theory testing as an assessment of the general validity of theories, in Bayesian terms when a case study with a very high prior probability of fitting a theory does not in fact fit the theory, this can greatly change our degree of belief in the theory in question (see Dion 1998, though Dion makes this point mostly with regard to a small number of cases, rather than a single case).

If we hold more deterministic assumptions about the world, for example, if we think single or conjunctive conditions of necessity or sufficiency hold, then (barring measurement error) a single case can disprove a claim of necessity or sufficiency. Munck argues that "deterministic" assumptions like necessity or sufficiency are "extremely implausible," but whatever one thinks of the plausibility of such assumptions, they are quite common: Gary Goertz has catalogued dozens of instances of social science theories that argue for necessity or sufficiency (in Goertz and Starr, 2002). At least for the many authors arguing such conditions, single-case studies can pose strong anomalies if measurement error is not to blame.

Perhaps more important and more common than these two types of theory testing in single-case studies, which work only in specified contexts, single-case studies are useful in narrowing or broadening the domain in which theories apply. As indicated above, I view theory testing as a question of specifying the proper domain of theories as much as it is a process of bolstering or weakening theories in general. Theories that fail a most-likely case need to have their domain reduced, and those that prove correct in a least-likely case, or a case in which all competing explanations predict a different process or outcome, deserve to have their scope widened (though Munck's reminder about the dangers of measurement error rightly qualifies the extent to which we should modify theories and their scope conditions as the result of the findings in a single case).

Finally, as Ronald Rogowski reminds us (in Brady and Collier, 2004), even if we are only studying a single case, we usually do so in the context of our background knowledge of the findings of other studies, both qualitative and statistical. Our assessment of the importance and generalizability of the findings of a single case are shaped in part by how those findings relate to this background knowledge. If the single-case study findings generally fit the pre-existing literature, they will not greatly revise our view, but if the case-study conclusions challenge the literature (and if we have confidence in the measures and methods of the study), we may greatly change our view of the existing literature.

6. Testing Causal Mechanisms, and the Fallacy of Distraction. There is perhaps no more profoundly and widely misunderstood dimension of case-study methods than their alleged susceptibility to a "degrees of freedom" problem. On the most general philosophical level, all methods suffer from a problem of *underdetermination*: we cannot know whether the theories we have are the best possible theories, since we can't compare them to theories we haven't thought of yet, and we cannot have definitive tests of theories as it is unclear whether a theory or its auxiliary assumptions has failed when a theory appears to fail to explain a case or a statistical distribution.

On a narrower level, specific methods suffer from particular kinds of *indeterminacy*, or inability to strongly and confidently infer which among competing theories that we have thought of best fits the evidence. In statistics, a particular form of indeterminacy is the degrees of freedom problem: it is impossible to carry out statistical analysis when one has more parameters than cases. In case study methods, however, whether a research design is indeterminate depends not on the number of cases and the number of variables, but on how the evidence from the case(s) stacks up against the contending theories. Case studies can present numerous observable implications on the processes that alternative theories predict should have taken place in the case, but whether these observations resolve the problem of indeterminacy is not a simple function of the number of observations and the number of theories. If the predicted implications that can be observed through process tracing do not differ between two contending and incompatible theories, there is no way to differentiate between them on the basis of the case, no matter how many observations one makes. On the other hand, there may be many contending theoretical explanations for a case, but a single piece of process-tracing evidence may be able to exclude all but one of them as an explanation for the case (as Stephen Van Evera suggests, think here of a bank security camera image that points to one suspect as guilty and exonerates all other suspects; Van Evera (1997) calls this a "doubly decisive" test. I concur here with Brady's and Collier's (2004) argument that process tracing on observations from within a case is quite a different method from statistically analyzing data-set observations across cases—each suffers from different limitations and different forms of potential indeterminacy.

7. Case-Oriented vs. Variable-Oriented Approaches and the False-Dilemma Fallacy. In my own methodological writings I have used the terms "variable" and "hypothesized causal mechanism" frequently, and I do not disagree with Munck on the usefulness of breaking down cases or processes into constituent parts under these labels. At the same time, I maintain that thinking about and studying cases in considerable contextual detail, or treating "cases as configurations" as many have put it, has advantages in identifying and incorporating both omitted variables and interaction effects. In a statistical analysis, if a researcher is coding his or her own cases, there is the potential for finding unexpected or previously un-theorized interaction effects within the cases. Once the cases are coded and the variables are set, however, the only interactions that can be found are those among the variables already specified (and even these may well be overlooked unless the researcher does a "specification search," which presents its own set of methodological challenges and pitfalls). Moreover, if a statistical analysis is based on pre-existing data sets, there is no process of potential discovery of omitted variables, and untheorized interaction effects can only be uncovered through error-prone specification searches. In contrast, when a researcher studies a case in great contextual detail, there is a much greater chance that s/he will inductively stumble upon variables and interaction effects s/he had not considered or theorized prior to looking at the case. There is no guarantee that the researcher will recognize all the important variables and interaction effects, but there is at least an opportunity to discover them in archives, interviews, and other sources.

8. Context and the Fallacy of Ad Hoc Explanations. Munck provides a useful warning against the dangers of ad hoc theorizing in case study explanations. There is indeed a risk of telling "just so" stories about cases when we study them in great detail. At the same time, a number of methodological writings, including my own with Alex George (2005), have urged that new variables uncovered from process tracing within a case should be systematically theorized, as Munck recommends. In addition, we argue that researchers should look for additional observable implications of the theory or explanation underlying the new variable, both within the case and across cases. Failure to do so does indeed raise the danger of curve fitting.

9. Causal Model Specification, Estimation Methods, and the Fallacy of Conflation. Rather than conflating the issues of (non)linearity and (non)determinism, I would argue that much of the recent literature has been quite clear on these issues. Goertz and Starr (2002) have explored both statistical and qualitative methods for addressing theories involving deterministic necessary or sufficient conditions. Ragin (2000) explicates methods for addressing theories involving probabilistic statements of "nearly necessary" or "almost sufficient" conditions. This does not necessarily commit either author to arguing that the social world is anywhere or everywhere deterministic or probabilistic (we will of course never know for sure due to the ever-present possibility of measurement error and model mis-specification). Rather, these authors are offering methods suitable for instances in which other researchers adopt or implicitly act upon deterministic or probabilistic theoretical assumptions in their empirical research. In these works and elsewhere (Goertz, forthcoming) these authors have been quite clear and sophisticated in addressing non-linear and non-additive theories, as well as in taking on the (separable) issue of (non)determinism.

10. Cumulation and the Questionable-Cause Fallacy. Cumulative knowledge about causal relationships is not only a matter of "summary measures of causal effects," "estimates of uncertainty and likely bias," "procedures to correct for selection bias," "tests of robustness," and "procedures to carry out a meta-analysis," as Munck suggests. These subjects are indeed useful to cumulative causal understandings, and statistical methods in observational settings, when their demanding requirements are met, can be quite useful at getting at estimates of these kinds of information. Qualitative methodologists, as noted above, have also devoted considerable efforts to addressing some of these issues, most notably the problem of selection bias.

Yet even if we had all this information with a high degree of confidence, we would not have the kind of cumulative knowledge about the social world that we desire. Alex George and I have written at length (2005) on why the notion of "causal effect" is only one part of explanatory causal theories, and why such theories must also make appeals to (ultimately unobservable) causal mechanisms. We have been equally critical of the idea that causal mechanisms are somehow prior to or superior to causal effects. Strong explanatory theories make arguments about both causal effects and causal mechanisms, and each kind of argument can have testable observable implications, some of which involve processes within cases and others of which involve comparisons across cases. Using process tracing to develop and test historical explanations of individual cases provides one kind of knowledge, closer to the notion of causal mechanisms than that of causal effects, that contributes to our understanding of the processes through which theorized variables exert their effects and the scope conditions under which these variables do and do not operate in specified ways. Cross-case comparisons, whether of a few cases by qualitative means or of many cases by statistical means, address other aspects of causal arguments.

Conclusion

Methodologies involve trade-offs among competing desiderata for effective cumulation of knowledge (Gerring, 2001). My argument is not that qualitative methods are invulnerable to many of the risks to which Munck has rightly drawn our attention. Rather, my argument has been that recent literature on qualitative methods has recognized and addressed, albeit imperfectly, many of the same problems Munck raises. In other instances, the recent literature does indeed differ from Munck in how it defines particular methodological problems or in how it assesses the trade-offs between methodological challenges as Munck defines them and other desiderata for good qualitative research. A clearer understanding of these tradeoffs is critical to continuing to improve qualitative methods, and Munck's essay makes an important contribution here by focusing our attention on fundamental issues.

References

Adcock, Robert and David Collier, "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research," *American Political Science Review* Vol. 95, No. 3 (September, 2001).

Brady, Henry, and David Collier, *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Rowman and Littlefield, 2004).

Collier, David and James Mahoney, "Insights and Pitfalls: Selection Bias in Qualitative Research," *World Politics* vol. 49, no. 1 (October, 1996).

Dion, Doug, "Evidence and Inference in the Comparative Case Study," *Comparative Politics*, Vol. 30, no. 2 (January, 1998).

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

MIT Press, 2005).
Gerring, John, *Social Science Methodology: A Criterial Framework* (Cambridge University Press, 2001).
Goertz, Gary, *Social Science Concepts: A User's Guide* (forthcoming).
Goertz, Gary, and Harvey Starr, eds., *Necessary Conditions: Theory, Methodology, and Applications* (Rowman and Littlefield, 2002).
Mahoney, James, "Nominal, Ordinal, and Narrative Appraisal in Macro Causal Analysis," *American Journal of Sociology*, Vol. 104, No.3 (January 1999).
Mahoney, James, and Gary Goertz, "The Possibility Principle: Choosing Negative Cases in Comparative Research," *American Political Science Review*, Vol. 98, no. 4 (November, 2004).
Munck, Gerardo, and Jay Verkuilen, "Conceptualizing and Measuring Democracy: Evaluating Alternative Indices," *Comparative Political Studies* (February 2002).
Ragin, Charles, *Fuzzy Set Social Science* (University of Chicago, 2000).
Van Evera, Stephen, *Guide to Methodology for Students of Political Science* (Ithaca: Cornell University Press, 1997).

How Qualitative Research Really Counts

Kenneth R. Benoit
Trinity College, University of Dublin
kbenoit@tcd.ie

> Polonius: What do you read, my lord?
> Hamlet (reading a book): Words, words, words.
> Polonius: What is the matter…that you read, my lord?
> Hamlet: Slanders, sir: for the satirical slave says here that old men have grey beards; that their faces are wrinkled; their eyes purging thick amber and plum-tree gum; and that they have a plentiful lack of wit, together with most weak hams: all which, sir, though I most powerfully and potently believe, yet I hold it not honesty to have it thus set down; for you yourself, sir, should be old as I am, if, like a crab, you could go backward.
> Polonius: Though this be madness, yet there is a method in't.
> (*Hamlet*, Act 2 Scene 2)

The main point of this essay is straightforward: The distinction between quantitative and qualitative research, when applied to empirical political analysis, is exaggerated and largely artificial. In fact, most political scientists can happily perform valid and useful research without being concerned about where they stand on the quantitative-qualitative divide. Furthermore, qualitative characterizations are often easily converted into quantitative characterizations, and many qualitative characterizations are implicitly quantitative to begin with. Finally, qualitative characterizations of the empirical world are almost always more useful when converted into quantitative ones.

In the spirit of a piece written for a newsletter on qualitative methods, I will at the outset fully acknowledge that my essay is a discourse and should be treated as such. As a bit of background on the context of this text and its author, I describe myself as a comparative politics scholar, primarily quantitative but also familiar with field work, interviewing, and survey analysis. I teach Advanced Quantitative Methods to Ph.D. students but also a course in Research Design. My current research involves estimating political party positions on policy issues in numerous countries and in the European parliament, using surveys of expert judgments and computerized content analysis.[1]

I will draw the distinction between quantitative and qualitative research in a deliberately simple manner, and then explore the implications of this distinction. The difference has to do with the use of numbers. *Quantitative* research characterizes observed phenomena using numbers, while *qualitative* research does not. A qualitative statement about voter attitudes toward political participation is that voters are disillusioned and apathetic, feeling that voting is a waste of time in the face of widely perceived corruption, ineffectiveness, and lack of meaningful policy content in party platforms. A quantitative statement would be that two-thirds of voters do not plan to vote, or that 45% report not feeling close to any particular party.

This simple distinction is normally confused by the unnecessary bundling of quantitative or qualitative research with other related yet logically separate issues. Such issues include the balance of cases to variables, whether research should be critical, normative, or positivist, and whether we can use case studies to prove causal propositions.

Let us deal with the first of these conflated issues, that the qualitative-quantitative distinction has to do with the number of cases, or more accurately, with the ratio of cases (call this n) to variables (call this k). Conventional accounts[2] of causal inference require that $n > k$, while qualitative researchers maintain that valuable knowledge, possibly even causal relationships, can be determined when n < *or* = k. Several points can be made on this issue. First, it is interesting that the identification of what is *qualitative* research in this framework rests on fundamentally *quantitative* grounds, namely the relationship of the quantities k and n. Second, to conceive of the qualitative-quantitative difference in terms of cases vs. variables makes it impossible to maintain that the two types of research are different in kind. Rather, it suggests that the difference is measured in degrees, even on a ratio scale, more precisely by the ratio of n to k. Finally, a focus on sample size shifts debate to other issues such as causal inference and case selection, obscuring the central issue of whether the empirical world consists of qualities or of things that can be counted.

Yet it is this issue of counting that is central in distinguishing quantitative from qualitative research. The essence of the matter boils down to measurement and the type of information we can feasibly use in characterizing the observed world. In the language of measurement, in fact, the distinction is more sophisticated than a simple dichotomous difference as implied by quality vs. quantity. Observations can also be measured according to different levels of *scale*, typically described as:

> *Nominal*: observations are distinguished from one an other in a purely qualitative fashion, such as parties, states, or ethnic groups.
>
> *Ordinal*: observations contain an inherent ranking, such as Fail, Poor, Good, Very Good, and Excellent. Ordinal measures are qualitative but can easily be converted into quantitative measures, such as assigning 0-4 to the previous example (perhaps to compute a grade-point average). This sort of conversion is also carried out by Likert scales, for instance.
>
> *Interval*: implies that observations can be measured on a scale where increments have a constant distance, such as when we measure temperature on the Fahrenheit scale. All interval scales are quantitative.
>
> *Ratio:* this is a purely quantitative scale that takes interval measurement further by having a meaningful zero point, permitting ratios to be taken. For instance, we might measure the number of cases in one's research design on a ratio scale, with 100 cases being 100 times greater than a single-case study, and zero cases representing only a theory with no data.

The move from qualitative to quantitative measurement occurs as more information is incorporated. It is also a natural consequence of any effort to *compare* observations. Comparison implies ordering, whether on a qualitative or quantitative dimension. Ordering implies by nature that one quality is stronger or greater in one observation than in another. And relations such as "stronger" or "greater" imply, whether this is made explicit or not, a relative degree of *quantity*, even if the characteristic being compared is discussed in purely qualitative terms. The act of comparison, therefore, naturally and readily lends itself to quantification.

I will take this reasoning a step further, to make a strong claim for the innate superiority of characterizations of the empirical world based on quantitative research. Our understanding of the empirical world rests on a system of statements supported through evidence. One of the primary objectives of empirical research is to establish this evidence. I contend that when it comes to establishing and defending such statements, quantitative evidence is superior to purely qualitative evidence. Evidence based on numbers is easier to compare, easier to verify, and easier to refute than that based on qualitative evidence. Even purely qualitative evidence, such as expert opinion, is elevated in reliability when it is expressed the implicitly quantitative framework of a consensus or experts.

Now at this point you may strongly disagree with my views, or you may disagree, or you may neither agree nor disagree, or you may even agree or strongly agree. I suggest that if you do not agree with any rating other than strongly disagree, then you have a logically inconsistent position if you call yourself a qualitative researcher. (A true qualitative position would permit only either categorical agreement or disagreement with the proposition that quantitative measurement is innately superior to purely qualitative representations.)

In the discussion to this point I have assumed that our enterprise was to characterize the empirical world. This returns to the second of the "conflated problems" I discussed above, which is the mode of inquiry. By restricting ourselves to characterizing the empirical world, we remove from the quantitative-qualitative research discussion not only formal theory, political philosophy, normative political argument, but also interpretative approaches such as discourse analysis, constructivism, social constructivism, post-positivist neo-feminist critical constructivism, and so on. These latter approaches share not only an inevitably strong qualitative element, but also a different basic objective from empirical ("positivist") research. At the extreme of these are interpretivist approaches which deconstruct reality as if it were a text, where the reader interacts with the text and its social context and attempts not just to uncover but also to construct meaning. Critical literary analysis, whether deconstructive or not, is typically interpretative, where the goal is to uncover meaning for the purpose of understanding a text, its story, the social world it represents, etc.

To draw on a more quotidian form of literary analysis consider film reviews. Film critics compare and evaluate, but with the goal of aiding the reader to understand and appreciate a film, in addition to knowing whether it is worth seeing. A reader of film reviews will typically know something about the critic's tastes based on a contextual knowledge of the critic's previous reviews, and will therefore be able to interpret the review accordingly. In this way, for instance, a reader of the *Times of London* might read between the lines of a one-star thumbs down from a culturally elitist British reviewer and, despite an unfavorable review, disregard the reviewer's suggestions and nonetheless go to see a perfectly good film like *X-Men*.

Text analysis is in fact an excellent field on which to pitch this battle, since text analysis involves fundamentally qualitative matter that may be analyzed either qualitatively or quantitatively. Let us assert, for instance, that George W. Bush is more of a conservative internationalist than a liberal internationalist like Woodrow Wilson. This statement about the empirical world may be considered an accurate characterization by many scholars of foreign policy. But ultimately such claims must rest on evidence. We might analyze a number of George Bush's speeches to provide this evidence. (Note the use of the term *number*.) Even if we only analyzed one speech, we could seek evidence in quantitative measures of certain words whose use would imply a particular foreign policy orientation. A key feature of conservative internationalism, according to Professor Henry Nau, is an emphasis on freedom over democracy. We might note then that Bush mentioned freedom 27 times in his inaugural address and 21 times in his State of the Union Address and not once stability.[3] This form of evidence is easier to compare—say to speeches by other U.S. presidents or other world leaders—easier to verify, and easier to refute, perhaps on the grounds that use of these words in these speeches is not an appropriate indicator of foreign policy orientation.

Table 1

Goal	*Quantitative*	*Qualitative*	*Examples*
Interpretation	No special value	Critical	Appreciate *Hamlet* Construct discourse on winking
Understanding	Useful	Critical	Understand *Hamlet* Understand meaning of winking
Description	Critical	Useful	Compare *Hamlet* to other Shakespeare plays Record observed winking ritual Describe a wine Award an Olympic gold-medal winner Characterize a country's democraticness Express or record a political preference Classify research as quantitative or qualitative
Explanation	Critical	No special value	Judge authenticity of *Hamlet* authorship Identify states likely to fail Determine whether campaign spending affects electoral success Determine the factors that influence ticket splitting

Consider another text. The preface to this essay quotes *Hamlet*.[4] The question of what to make of this text is not unlike the question facing researchers confronted with the political world. Are we concerned with whether Hamlet is really mad or merely faking it? What should we conclude about the book Hamlet is reading, based on Hamlet's description? Are we to understand that in his oblique invention of a text about nasty old men in his reply to Polonius, and in mistaking Polonius for a lowly fishmonger just before this exchange, Hamlet is violating strongly held cultural norms regarding politeness and the display of respect for one's elders?

Those are goals of interpretation. We might also attempt to establish falsifiable, empirical statements about Shakespeare's texts. For instance, we could examine *Hamlet* as a whole to determine stylistic evolutions between this and later plays written by Shakespeare. Or, we might attempt to determine the authenticity of authorship based on cryptographic clues possibly left by Frances Bacon or some other ghost-writing impersonator. As it turns out, such debates actually occupy a great deal of space in the literature on Shakespearean literature.

An example: In 1985 a new poem was discovered in the Bodlean Library of Oxford by Gary Taylor and attributed to Shakespeare. On what evidentiary basis would we consider it authentic? Authoritative declarations from Oxford Shakespeare experts? Considered more definitive was quantitative evidence established by two statisticians at Stanford University, Bradley Efron and Ronald Thisted, who statistically analyzed the Bard's entire 900,000-word vocabulary in order to establish usage patterns. Efron and Thisted tested this distribution against the writings of Shakespeare's contemporaries, using the rules of statistical inference, and found that the test distinguished clearly the writing of Shakespeare from Donne, Marlowe, and Jonson. They then used the observed distribution to predict similar patterns in the 430-word mystery poem and concluded that it perfectly fit the profile for Shakespeare's work.[5]

In having brought literary analysis into this discussion at all, however, we find ourselves at the brink of a dangerous pit regarding the notion of science and whether the author or investigator does or should matter in research, and whether political science research should be closer to the stereotype of hard science or whether it can or should share elements of film reviews.

I am going to go circumvent this pit, however, by arguing that even film reviews and interpretive accounts can be enhanced by using quantitative information. For film reviews, a "thumbs up" or "thumbs down" may be qualitative, but a rating from one to five stars[6] converts the qualitative, highly subjective, interpretative measure into quantitative information. Wine and cigar reviews do the same, such as not only describing the Carlos Toraño Signature Collection as having an "earthy core with hints of leather and sweet spice," but also as having received the ("astounding and nearly unprecedented") rating of 4.7 from *Smoke Magazine*.[7]

In the summer of 2004 we probably all watched some of the summer Olympic events. Judging Olympic events involves making highly subjective judgments of observable behavior, where the past experience, artistic and athletic context, and personal orientation of the judges all play an important and acknowledged role. Many competition events have formal guidelines (such as the mandatory components of gymnastic routines), but winners are determined ultimately according to intangible qualities. In other words, the excellence of a performance is an intrinsically qualitative characteristic. But once we need to *compare* performances, such as determining three top-ranked winners, then this qualitative performance *must* be measured quantitatively. Of all the events in the Olympics, there is not a single event—whether based on solo artistic or technical performance, or on scoring goals—whose outcome is not determined on the basis of a quantitative score.

So, to summarize, in order to enhance their usefulness qualitative measurements are generally converted into quantitative information. We need reach no further than our bookshelf or latest journal copy to find abundant examples from our own discipline: location of political party positions on a left-right scale; public attitudes toward post-materialist values; levels of democratic governance or corruption; the levels of conflict in international environment; the relationship between electoral systems and the expected number of political parties. Goals that we value in political science, such as the ability to make meaningful comparisons, the manageability of data, the ability to replicate analyses, the capacity to characterize confidence or uncertainty, and the potential of our propositions to be falsified are all enhanced when research rests on quantitative evidence. Purely qualitative approaches are most useful if we wish to interpret or understand an observed phenomenon (event, idea, text, etc.) but if we wish to compare interpretations, we are likely to need numbers. (See Table 1above for an attempt I have made at classification in order to maximize comparison and potential refutability of my claims.) Purely qualitative description is possible, but comparison implies relative quantities, meaning that this form of qualitative research really *counts*. Explanations of the sort in which we are likely to have confidence will involve some form of quantitative statements 100% of the time.

Notes

[1] A description of and data from the expert survey project is available from http://www.politics.tcd.ie/ppmd/. A full description, research papers, and software for the computerized content analysis project are available from http://www.politics.tcd.ie/wordscores/.

[2] See Gary King, Robert Keohane, and Sidney Verba (1994) *Designing Social Inquiry* (Princeton University Press), known in many circles simply as "KKV."

[3] Henry R. Nau, "Bush's classic conservatism." *International Herald Tribune*, March 29, 2005, p7.

[4] William Shakespeare, 1603. *Hamlet* is a play whose namesake is a Wolverine-type character (*X-Men*) in the sense that the has parental issues, those close to him worry about his stability, and he inadvertently stabs a good guy. (And on what basis can you judge this literary interpretation worse than any other?)

[5] Bunce, Nigel, and Jim Hunt, "The Statistics of Shakespeare," The Science Corner, May 13, 1968, University of Guelph, http://www.physics.uoguelph.ca/summer/scor/articles/scor109.htm.

[6] Or tomatoes; see for instance http://www.rotten-tomatoes.com.

[7] See https://www.cigarsforless.com/Cigars_in_US/Carlos_Torano_Signature.htm for the full description and ordering details.

The American Political Methodology Debate: Where is the Battlefield?

Bernhard Kittel
University of Amsterdam
b.e.a.kittel@uva.nl

Like many other realms of political science, the debate over political methodology in Europe has been influenced by the American discussion up to the point of being reduced to a perhaps somewhat belated commentary of debates deemed terminated in the American context. An analysis of references to methodological contributions in political science articles would probably reveal that American authors are far more frequently cited by Europeans than vice versa. It is even hard to discern something like a European methodology debate worth of its name.

Recent observers of the differences between American and European practice have in particular highlighted the stronger emphasis on and disciplinary status of comparative politics in Europe as compared to the United States (Lijphart 1997) and the more systematic empirical-analytical approach using quantitative data for rigorously testing hypotheses in the United States as compared to more institutionalist, descriptive, constructivist, and more generally qualitative approaches in Europe (Marsh and Savigny 2004; Moses, Rihoux, and Kittel 2005; Norris 1997). These stereotypes, which can already be found in David Lodge's characters of Philip Swallow, the worrisome British academic, and Professor Zapp, the jovial American versed in the ways of the world, in his novel *Changing Places*, contain, like all of such generalizations, some elements of truth. But reality is always more complex and we can find practitioners of all denominations in both academic communities.

Looked at from a more long-term European perspective, the current American debate reiterates episodes which we have encountered in the European history of science. The battle between the nomothetic and the idiographic worldview has accompanied the social sciences since their first attempts to define their topics and approaches, and the relationship has never been one of great friendship. Among these, the debates between social philosophy and positivist social science in the early 19th century, the economic *Methodenstreit* of the late 19th century between the Austrian marginalist school and the German historical school, and the *Positivismusstreit* waging between the critical theorists of the Frankfurt school and the "positivists" during the 1960s are only the most notable. In this perspective, the current American debate may simply appear as the newest clash of academic civilizations. In comparison to these older debates, however, the current controversy is indeed astonishing in the extent to which the contending proponents seem to converge on fundamental issues. In this sense, perhaps, we could speak of a very American solution to the long-standing conflict.

The American debate has induced an unprecedented awareness of methodological problems and fallacies, and its impact on the world of political science has certainly been to make it a better science. Notwithstanding the respect due to all contributors, and without claiming undue originality, I would like to raise three contentious issues which seem to be regarded as uncontroversial in the American debate. These concern the cumulation of research, the generalizability of macrophenomena, and the notion of probabilistic causality. My suggestion is that by accepting the quantitative template in these respects, qualitative analysts unneccesarily weaken their case. In contrast, I think that in the longer run also quantitative research may profit from a more critical position on these issues by qualitative researchers.

Cumulation

It is a truism that cumulation of knowledge is an important goal, not only of scientific research but of human life in general. We link different pieces of knowledge in order to obtain a broader understanding of our environment. This may be accomplished intuitively or systematically (scientifically). And we do so when we believe situations to be sufficiently complementary for linkages to be valid.

What is, then, surprising about the emphasis that the current discussion places on cumulation? King, Keohane, and Verba rather uncontroversially recommend to "contribute to a scholarly literature by increasing the collective ability to construct verified scientific explanations of some aspect of the world" (King, Keohane, and Verba 1994: 16-17). Granato and Scioli (2004: 314) specify this recommendation as implying building models of social reality, deriving hypotheses from these models, and testing them on data. They explicitly reject any difference between formal and informal modeling approaches and between quantitative and qualitative data analysis in their ability to contribute to the cumulative extension of our knowledge. And, from the qualitative perspective, Mahoney and Rueschemeyer (2003: 23) add: "In contrast to interpretive and postmodern analysts, comparative-historical researchers defend the enterprise of causal analysis, though they do not aim for the universalistic kind sought by some rational choice theorists." Thus, aside from some family quarrels at the level of analytical approaches, there seems to be widespread consensus about the aims of doing political science.

I have no objections to the general goal of cumulation. But we should harbor no illusions about the degree to which this goal is attainable. Given this reality, it becomes problematic if the cumulation-enhancing potential of research projects is made a core criterion for institutional support. In my view, this criterion is as crippling for ultimate scientific progress as the "proliferation of noncumulative studies" (Granato and Scioli 2004: 313).[1] Implicitly, the standards of the two dominant disciplines in the academic world, physics and economics, are held up as exemplary paradigms for the rest of us. By building encompassing, logically consistent, theoretical foundations, these disciplines are said to provide models of natural and social reality which should allow us to develop guidelines for model-building, intervention, and manipulation in other realms.

But such a view of physics is not shared by the practitioners of modern physics themselves. Most serious physical scientists now accept both the particle and wave theory of light as different ways of conceptualizing that aspect of the natural world. They also accept the probabilistic nature of subatomic events in quantum theory, the uncertainty principle, the impossibility of distinguishing between uniform accelerated motion and a gravitational field, the failure of local causes, the concept of reverse time, the impossibility of refuting the assumption of the existence of parallel universes, and other mind-boggling evidence of noncumulation.[2] More generally, in his famous theorem Gödel has shown that there exists no consistent set of statements that can ever hope to deal with all possible propositions (Hofstadter 1979). There is no apparent reason why this theorem should not apply to theoretical or modeling endeavours outside of mathematics.

The failures of unification in economics, which has proceeded far in building an axiomatic, internally consistent edifice, are even more apparent. Sure, economic models are to some extent able to predict outcomes. But these predictions are based on rather restrictive assumptions, among which the identity and independence of a large number of individuals under homogeneous conditions and the utility maximization axiom are noteworthy. This severely limits the set of real-world situations that can be accurately captured by economic models. And why are there still so many negative externalities of economic policies based on the collected knowledge of the world's economists? While the certitudes of economics are gradually being undermined, at least at the philosophical fringe (Bunge 1998; Fullbrook 2003; Mäki 2002; Rosenberg 1992), political science appears to be mimicking the errant adventures of its much-admired big brother.

If the above reasoning is accepted, its implications for the project of disciplinary cumulation and eventual unification are disturbing. Indeed, such an attempt, in my view, is bound to fail. It also flies in the face of the Popperian principles of conjectures and refutations, which invite an eternal scepticism against whatever certitudes we might decide to embrace. Stated even more bluntly, if the likes of Kepler, Galilei, Newton, Darwin, or Einstein had adhered to King, Keohane, and Verba's or Granato and Scioli's calls for discipline, their signal contributions to scientific progress may well never have been attempted. There is currently no scientific discipline which can reasonably claim to have a TOE (Theory of Everything), not even in its own field. Instead, the scientific landscape looks much more like a "dappled world" (Cartwright 1999) in which small illuminated islands of knowledge present a partial and incomplete picture of reality without giving clues as to how much is not known and how the observed pieces might connect to each other.

Generalizability and Theory Testing

In my view the methodological debate between "quants" and "quals" which we currently witness in American political science is much less fundamental than its contributors would have us believe. Specifically, the approach taken by the qualiitative side (Brady and Collier 2004) surrenders too much. Most notably, Collier, Brady, and Seawright (2004: 244-249) reduce the differences between quantitative and qualitative research to issues of levels of measurement, numbers of cases, the use of rigorous (statistical) tests, and "thick" versus "thin" analysis. By thus accepting the quantitative ontology and epistemology, the true underlying ambivalence of our conception of social reality is hidden from view.

I will attempt to substantiate this point by elaborating on the second issue, one that, from a European perspective, is somewhat surprising. In contrast with early contributions to the "comparative approach" (Smelser 1976; Tilly 1984), recent work in the field of comparative-historical analysis seems to have become obsessed with theory testing and generalizability. According to a state-of-the-art summary, comparative-historical analysis is defined as "a field of research characterized by the use of systematic comparison and the analysis of processes over time to explain large-scale outcomes such as revolutions, political regimes, and welfare states" (Mahoney 2004: 81, fn. 1).[3] Moreover, it is stated that "all work in this tradition does share a concern with causal analysis, an emphasis on processes over time, and the use of systematic and contextualized comparison" (Mahoney and Rueschemeyer 2003: 10). Implicitly, the emphasis on causal analysis embraces the concept of causal laws (Mahoney and Rueschemeyer 2003: 22), which are supposed to be ascertained by comparing similar social macrophenomena. A number of tools have been developed to help qualitative researchers to identify laws or test hypotheses derived from such laws, ranging from the famous and often misunderstood most similar/most different system designs (Przeworski and Teune 1970) via a resuscitation of Mill's methods (Skocpol and Somers 1980)[4] and various principles of theory-driven case selection (Geddes 2003; Goertz and Starr 2003), to the Boolean algebra of Qualitative Comparative Analysis and its derivates (Ragin 2000).

My impression is that in these attempts to deal with King, Keohane, and Verba's definition of research standards, comparative-historical analysts have manoeuvred themselves into a trilemma. As far as I understand the above elaboration, they attempt to answer general questions about the *operation* of social phenomena by analysing the *origin* of specific social events at the level of social *macro*phenomena. The claim which I would like to defend here is that this specific combination of these three elements leads comparative historical analysis into an impasse.

The distinction between queries into the explanation of the *operation* of something and the explanation of the *origin* of something has been noted variously. For example, Dessler (2003: 22) highlights the difference between the search for causal generalizations, which he identifies as "positivist social science in its modern form," and the "reconstruction of detailed historical episodes," which, according to him, is the realm of interpretivism. In Casti's (1989: 173) formulation, which primarily refers to the natural sciences, an essentially identical distinction is made between *operation science*, which aims at explaining repeated (and repeatable) *processes* in a system, and *origin science*, which is concerned with the explanation of singular *events*. Examples of the former are the passage of the Earth around the sun, the union of hydrogen and oxygen to form water, or the flow of electrons through a resistor. Among the examples he cites for the latter are the formation of the universe, World War II, the composition of the Mona Lisa, and the origin of life. Or, to cite a well-known philosopher of science: "When we speak of scientific explanations…[f]or the most part we have in mind explanations of *why certain phenomena occur*. The phenomena may be particular facts or general regularities" (Salmon 1998: 6, emphasis added).

The "operation mode" of analysis which focuses on the explanation of general regularities is very much what realist philosophers of science seem to refer to as the analysis of mechanisms in order to provide an answer to the question "how does it work" by tracing the way the components of a system interact. A system is here defined as "a complex object whose parts or components are held together by bonds of some kind" (Bunge 2004: 188). Hence the operation mode always shifts from the macro level of the system to the micro level of the components in order to study the mechanisms of the system which are defined as "processes in concrete material systems" (Bunge 2004: 191). The important ontological assumptions here are the systemic aspect of the phenomenon and the repeatability of processes. Systemic qualities allow the researcher to identify a border between the environment and the object to be studied within which different relations between its components–structures in which the processes take place–correspond to distinguishable but comparable states of the system. Repeatability of the processes implies that the system exhibits some persistence over time and that systems of the same kind will exhibit the same mode of operation.

I wish to draw attention to three problems involved in the apparent consensus in the American debate with regard to this difference. The first problem arises as a result of the sort of phenomena typically analysed by practitioners of comparative-historical analysis. In my reading, the modal focus of this approach is not in explaining the operation of systems but in explaining singular events or particular facts: revolutions, regime changes, welfare retrenchment, and events like that.[5] Such phenomena are neither systems nor repeatable. Moreover, they cannot be studied by asking "how do they work?" Rather, we must ask "how and why did they come about?" Thus, they call forth a different type of explanation. By accepting King, Keohane, and Verba's conception of research standards, analysts squeeze their work into an inadequate framework because it starts from the premise that a science must establish causal laws about the operation of a system. But when particular facts are the focus of analysis, a causal explanation does not establish general regularities but instead makes use of known regularities for providing an account of the origin of the observed singular event.

This leads to the second problem which, in my view, is hidden in the very conception of what constitutes a case. Ragin (1992) has suggested a very useful heuristic by distinguishing between specific and general conceptions of cases, on the one hand, and between an understanding of cases as empirical units and theoretical constructs on the other. While most quantitative research resorts to a general conception of empirical units, like individuals or nation states, comparative historical research tends to focus on empirical realizations of an ideal-typical concept, which uses an understanding of a case as a theoretical construct. As such, it is perfectly legitimate to identify, for example, a regime change in a particular situation as a case worthy of being studied. But the case is then defined by the researcher as being an instance of some social macrophenomenon defined by him/herself theoretically. Cases conceptualized in this way are not systems in the sense defined above but events. Therefore, such a case study must be a study of the origin of a particular fact, the regime change. The inconsistency comes in when concepts for analyzing the operation of systems are introduced into the analysis of the origin of events. For example, Gerring (2004: 342) uncontroversially defines a case study as "an intensive study of a single unit for the purpose of understanding a larger class of similar units." But further on it becomes clear that his conception of units is general-empirical, as he uses as examples "the person, group, organization, county, region, country, or other *bounded phenomenon*" (Gerring 2004: 344, emphasis added). But comparative-historical analysts claim not to be interested in the operation of France or Germany, but rather in the determinants of, for example, regime changes. Events happening in these countries are just realizations of ideal-typical social macro-phenomena in real-world political systems. This tacit shift from events to systems as the constitutive elements of a study forces a framework suitable for the analysis of the operation of systems onto a type of research question focusing on the origin of events. Moreover, what could be reasonably be conceived of as proper units of analysis in a political world characterized by complex governance, cross-cutting agencies (e.g., NGOs), and outcomes not unambiguously attributable to particular units (e.g., global warming)? This does not mean that we cannot study social revolutions, regime changes, or welfare state retrenchment. But I see little prospect in trying to establish "causal laws" operating at the macro level by means of comparing "cases" of such phenomena when the cases for which the laws allegedly hold are the analyst's own constructions instead of being constituted by social practice.

This last point introduces a third problem. In comparative-historical analysis, cases are defined at the macro level of social phenomena. One of the major points of disagreement be-tween quantitative and qualitative methodologists in the ongoing debate is the problem of homogeneity of cases in macrocomparative research. It has been a prominent and important critique forwarded by the qualitative side against typical studies in the quantitative tradition (Adcock and Collier 2001; Van Deth 1998), although it remains an open question whether any of the specific fixes offered–like, for example, context-specific indicators–will prove workable in practice. Both the contextuality and the historicity of the phenomena which qualitative researchers typically analyse contradict the possibility of deriving general statements which, by definition, should be essentially independent from time/space coordinates. None of the "cases" identified in macrocomparative research, be it quantitative or qualitative, plausibly fit this criterion. Independent nation-states cannot be regarded as relevant units outside of a particular epoch in history and outside of a particular portion of the world. And even in this particular situation, for which it is routinely claimed to hold, the assumption of independence is heroic indeed. Hence, the population of cases to which any such generalization from a particular comparative analysis should be made remains fairly opaque. Neither can we ever hope even to define a population of events of a certain class or to assume that Tilly's (1984) big structures and large processes occur independently of each other, which is a precondition for our ability to identify laws. In the presentation by Collier, Brady and Seawright (2004), "thick analysis" is meant to take into account nuances and complexities which measurement in quantitative approaches cannot attain. But these nuances and complexities simply highlight the heterogeneity of cases, which ends up invalidating the result. In this regard, I read Yanow's (2003) intervention as expressing the concern that the strengths of case studies might be given up as soon as they are reduced to "instances of something."

This suggests that the problem is broader than "merely" one of heterogeneity and lack of independence. In the social sciences, we tend to conceptualize such phenomena in terms of ideal types, mental constructs based on the one-sided accentuation of traits of a situation which the analyst believes to be relevant. The identification of cases of some ideal-typical macrophenomenon is the very foundation of the comparative-historical approach. However, this procedure is based on a literal interpretation of Weber's (1949) early, "holistic" conception of ideal types. Watkins (1952) regards the combination of holistic ideal types with the method of historical analysis, which is exactly the road taken by the scholars working in the comparative-historical tradition, as the core weakness of Weber's approach. By moving from the broad characteristics of an entire historical situation to particular deviations, the approach fails to take into account of the fact that "if knowledge of the general characteristics of a social situation is always derivative knowledge, pieced together from what is known of individual's situations," a procedure moving in the reverse direction from the general characteristics to the individual situation is logically impossible, because "the former is logically derivative from the latter" (Watkins 1952: 27).

Watkins conjectures that Weber (1978), in his later work, has tacitly abandoned this conception in favor of an individualistic approach for exactly that reason. "An understanding of a complex social situation is always derived from a knowledge of the dispositions, beliefs, and relationships of individuals. Its overt characteristics may be *established* empirically, but they are only *explained* by being shown to be the resultants of individual activities" (Watkins 1952: 29). In consequence, according to this interpretation of Weber's view, individualistic ideal types are constructed by "placing hypothetical, rational actors in some simplified situation, and in deducing the consequences of their interaction" (Watkins 1952: 29). Here we are, on the one hand, back at the realists' explanatory strategy via mechanisms and, on the other hand, the rational choice approach to comparative politics (Levi

1997). Within this broad framework, rational-choice predictions may be regarded as counterfactuals which provide a reference point for the analysis of empirical phenomena (Scharpf 2000). I will come back to this idea below.

Why does the foregoing constitute a problem for macro-comparative analysis? In my view this approach searches for law-like regularities where the preconditions for the existence of regularities do not exist. From this perspective, the enterprise of causal analysis via the comparison of macro-phenomena is an inherently problematic endeavour for both the operation and origin mode of analysis, because it has to come to terms with contingent associations, complex interdependencies and process interferences at the level of the units of analysis (Mayntz 2002). The complexity of macrosocial phenomena is simply too large to deal with in terms of this quantitative template: "Certain phenomena that macrosociologists have sought to study… turn out to be ones on which theory can give relatively little cognitive grasp at all… [T]hey appear too few, too independent, and too causally heterogeneous for anything of much use to be said in theoretical terms" (Goldthorpe 2000b: 63). Similar reservations have been expressed, inter alia, by Little (2000), Lieberson (1997) and Hall (2003).

Determinism and Probabilism

A third issue, closely related to the problem of generalizability, is the juxtaposition of determinism and probabilism in inferential statements. Bennett (1999) has rightly criticized King, Keohane, and Verba's conception of probabilistic statements and drawn our attention to the distinction between the inherent probabilism in the physical sciences' concept of quantum phenomena and probabilistic statements arising from our lack of knowledge about the world, which he terms "complexity-induced probabilism." I would like to take this point one step further by submitting that only the former conception is defensible. Both conceptions carry a concept of determinism in their back-bag. According to quantum logic, we can expect deterministic behavior at the macro level as long as a sufficiently large number of micro units are considered (Salmon 1998: 25-49; see also Waldner 2002). We do not know what will happen to any specific micro unit but provided that the model is adequate and the rules of statistical aggregation hold (identity and independence of behavior), we can predict with certainty what will happen at the aggregate level. Examples of this are the laws of thermodynamics, where molecular micro-behavior could in principle be analysed, and the half-life periods of radioactive decay, where the quantum processes cannot be traced in principle. Hence, despite a probabilistic conception of micro-behavior, we can make deterministic statements about macro-behavior. These statements are deterministic inferences derived from an ontological probabilism and it is apparent that determinism enters as a stylized description–a model–of reality, not as an assumption about reality. In other words, it is the model which is assumed to be deterministic, not reality.

Complexity-induced probabilism, instead, starts from the ontological presumption that the real world is deterministic. It is assumed that there is a "true" model out there, and any failure to correctly predict an outcome is considered as being due to omitted variables, measurement error, random events, or other contingent nuisances of the research process. Hence this conception results in probabilistic inferences constructed upon ontological determinism. According to complexity-induced probabilism, events occur only with a certain probability because we fail to produce a sufficiently detailed model of the world. This view, however, fails to adequately distinguish between theoretical and empirical models on the one hand, and between probability statements based on risk and those based on uncertainty on the other.

First, theoretical models are based on assumptions which simplify the problem and allow it to be treated as a closed system consisting of logical or mathematical equations from which predictions are derived (Sayer 1984: 165-180). In empirical applications of such models, probabilism enters in empirical applications as a tool for judging actual deviations from the expectations generated by the model due to random sample biases. In contrast, empirical models are open–indeterministic –systems which deal with various parameters, ranging from uninterpreted constants to the error term, in terms of unknown contingencies. Such open systems are inherently incapable of producing regularity statements from which expectations can be deduced, because any indeterministic model can be fitted to the data. Using such models in macrocomparative research implies that the probabilistic character of the model, represented by the error term, is derived from the relative frequencies in empirical observations without the sample being *randomly* selected from a *known* population. Claims like the one that a particular condition "is necessary or sufficient 90% of the time" (Mahoney 2004: 84), which is a statement based on a frequentist conception of probabil-ity, depend on the asymptotic properties of large random samples, which are unavailable in macrocomparative research.

Second, while risk-based statements refer to statistical aggregates which themselves are based on a known probability distribution in a circumscribed population of cases, uncertainty arises if we cannot assign a meaningful probability distribution to outcomes because we do not have complete knowledge of the possible outcomes and their probabilities. In this situation, any probability statement is nothing more than an expression of faith. In substantive terms, neoclassical economists have invested much effort in reframing uncertainty in terms of risk, and this seems to be precisely the reason why many political scientists and sociologists reject rational choice theory (Beckert 1996; Mahoney and Rueschemeyer 2003).

In methodological terms, we cannot attribute objective probabilities to singular events because there is no meaningful conception of a population for which the cases we analyze are representative and from which we can derive long-run propensities (Gillies 2000). In spite of paramount claims to the contrary in the quantitative literature, I do not accept that the concept of a random sample is meaningful at the level of social macrophenomena, regardless of how the units are defined. While the natural sciences now seem to base their inferences on the quantum-probabilisitic conception of the world, the quantitative template in the social sciences apparently adheres to the concept of complexity-induced probabilism. As far as I understand their work, the branch of the qualitative response to the quantitative view which aims at developing probabilistic statements about macrophenomena based on subjective probability assessments and Bayesian reasoning (Dion 1997) seems to take this conception for granted. I think that two arguments militate against this view.

First, in contrast to the class of human beings themselves, structured aggregates of human beings vary in space and time to an extent which makes it hard to believe claims of generalizability beyond a small set of strongly interrelated cases during a small period in time. Contemporary nation-states are particular instances in a process of constant organization and reorganization of human collectivities, self-selected by historical or political processes and constructed by historical contingency (Ebbinghaus 2005). For example, even within the small set of European countries since 1945, which comparativists believe to be the set of countries which is the closest real-world approximation to similarity, we have witnessed boundary changes (e.g., Germany, former Yogoslavia), far-reaching constitutional changes (France, Central European Countries), and the partial demise of formal independence across cases (European Monetary Union). In my view, as a basis for testing theories, this is closer to a comparativist's nightmare than to a sound basis for setting up a comparative design of independent units.

Second, the presumptions underlying the possibility of statistical aggregation from the micro to the macro level do not hold for many phenomena. As soon as we move beyond the statistical aggregation of individual behavior to the behaviour of macro-level units, i.e., socially structured phenomena, there is no way to tell whether two phenomena are sufficiently similar to warrant putting them into the same category. This holds both for the conception of cases as systems and cases as events, but the problem may be even more severe for the latter. While a social system still is based on some notion of boundedness and we might, for example, refer to the class of all currently existing sovereign countries or all members of the United Nations, it is much more difficult to establish boundaries for events, and hence it is impossible to conceptualize an unambiguous notion of events. This implies that there is no basis for defining a population which could provide the frame of reference for a statement on the long-run propensity of an event to occur. We are thus left with uncertainty, which results in a type of causal association, for which Bennett (1999) has found the telling notion of "enigmatic" causality. This notion–quite rightly–surrenders the issue of causal explanation to the realm of detective stories.

Conclusion

There is an alternative, to which various contemporary streams in the social sciences seem to be gradually converging (Abell 2003; Archer 1995; Bates et al. 1998; Coleman 1990; George and Bennett 2005; Goldthorpe 2000a). In order to establish an explanation of social macrophenomena, we can make use of observed regularities at the micro-level of individual behavior, which we then put together in recounting the story of this or that macrophenomenon. That is, we build an explanatory story by referring to a sequence of regularity statements at the microlevel which are connected by narrative elements. Thus, we explain regularities or singular events at the macro-level by elaborating the causal mechanisms relying on regularities at the micro-level. This strategy can be traced back to Max Weber (perhaps further). In philosophy of science, it now has gained widespread acceptance beyond the various streams of realism, which itself seems to have replaced positivism and critical rationalism as the dominant paradigm of scientific practice—outside of mainstream quantitative social science. Outstanding proponents of the comparative-historical school, among them Mahoney (2004), have shown great interest in "mechanismic" explanation strategies. So why don't these authors draw the conclusion that the search for generalities and the theory-testing endeavours work neither in quantitative nor in qualitative analyses of social macrophenomena and that regularities found at that level can at best provide puzzles to be explained by referring to probabilistic regularities at the micro-level?

To summarize my argument, my core objection to the position taken by the proponents of qualitative methods in the current debate is that they yield too much at the outset, because despite their critique of the quantitative template, they seem to have decided to fight the war on its premises. In contrast, I believe that, besides the many important elements qualitative researchers have quite rightly taken over from their quantitative peers on the level of methodology, the quantitative orthodoxy leaves too many open questions at the ontological and epistemological level of social science research to be accepted without reservations.

I should stress that I do not favor the outright rejection of cumulation, theory-testing, and generalization. My concern is that the type of phenomena typically addressed by comparative-historical analysts does not lend itself to generalizations or law-like statements. Instead, I argue that we have to take seriously the view that all we can do is to try to make sense of what we experience by developing models of parts of the world in the knowledge that our perception is based on our socially constructed conception of the world, and hence that what we see is what we know. I have highlighted three issues, the problematic nature of which seems to have been disregarded in the current American debate: the preoccupation with cumulation, the ambition of theory-testing and generalizability, and the notion of probabilistic causation are all embedded in a rather naïve, positivistic conception of social macrophenomena. In these respects, the qualitative response to the quantitative template seems to have embraced too many assumptions, and done so too readily. The acceptance of these assumptions blatantly disregards important developments both in the natural sciences and in the philosophy of science. Hence, the practitioners of comparative-historical analysis should not cripple themselves with ambitions to meet unattainable standards but, rather, should orient their energy toward defining standards of external validity that are more appropriate to their endeavour. I therefore believe that the discussion about shared standards should be focused less on the level of methodology and more on the level of epistemology. As such, the basic approach of comparative historical analysis should be rephrased in order to overcome its trilemma: A social *macro*phenomenon might be explained by tracing the causal mechanisms underlying its *origination* which rely on the *operation* of social *micro*-phenomena.

Notes

[1] A surprising aspect of the cited paper is that the authors deplore the lack of cumulative research but then elaborate on unrealistic modeling, potted histories, and sloppy statistics. These problems have nothing to do with noncumulativeness of research and everything to do with a lack of scientific rigor.

[2] I do not claim competence in physics beyond the high school level. My reading mainly stems from the excellent introductions by Zukav (1979) and Casti (1989).

[3] Here, and in the following, I use this paper as a pars pro toto because of its virtue of being an excellent and concise overview of the issues pertinent in comparative-historical analysis.

[4] The reverence paid to Mill's methods of agreement and difference among qualitative comparativists despite their both naive positivist and inductivist stance (Blaikie 1993: 135-138) is one of the strangest inconsistencies, which I can only imterpret as a tribute to the hegemony of the quantitative template. It is now well known that Mill himself did not believe his method suitable for the social sciences, essentially because we never know whether there is another omitted variable lurking behind the next tree of knowledge.

[5] I have taken the liberty to specify Mahoney's (2004: 81, fn. 1) formulation in order to stress the event character of the topics studied. In the area which I know best among those cited, welfare states, comparative-historical analyses have typically contributed studies of particular events and series of events, i.e., developments (Bonoli 2000; Huber and Stephens 2001; Pierson 1994). I am not sure whether this interpretation does justice to Mahoney's aims. But in my reading his statement seems to confound phenomena which can be conceived of as systems (political regimes and welfare states) and phenomena which cannot (revolutions).

References

Abell, Peter. 2003. On the Prospects for a Unified Social Science: Economics and Sociology. *Socio-Economic Review* 1 (1):1-26.

Adcock, Robert, and David Collier. 2001. Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review* 95 (3):529-546.

Archer, Margaret S. 1995. *Realist Social Theory: The Morphoge netic Approach*. Cambridge: Cambridge University Press.

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

Beckert, Jens. 1996. What is Sociological about Economic Sociology? Uncertainty and the Embeddedness of Economic Action. *Theory and Society* 25:803-840.

Bennett, Andrew. 1999. *Causal Inference in Case Studies: From Mill's Methods to Causal Mechanisms*.

Blaikie, Norman. 1993. *Approaches to Social Inquiry*. Cambridge: Polity Press.

Bonoli, Giuliano. 2000. *The Politics of Pension Reform. Institutions and Policy Change in Western Europe*. Cambridge: Cambridge University Press.

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

Bunge, Mario. 1998. *Social Science under Debate*. Toronto: University of Toronto Press.

Bunge, Mario. 2004. How Does It Work? The Search for Explanatory Mechanisms. *Philosophy of the Social Science* 34 (2):182-210.

Cartwright, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Casti, John L. 1989. *Paradigms Lost: Images of Man in the Mirror of Science*. New York: Morrow.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge: Belknap.

Collier, David, Henry E. Brady, and Jason Seawright. 2004. Sources of Leverage in Causal Inference: Toward an Alterna tive View of Methodology. In *Rethinking Social Inquiry. Diverse Tools, Shared Standards*, edited by H. E. Brady and D. Collier. Boulder: Rowman and Littlefield.

Dessler, David. 2003. The Positivist-Interpretivist Controversy. *Qualitative Methods* 1 (2):21-24.

Deth, Jan van, ed. 1998. *Comparative Politics: The Problem of Equivalence*. New York: Routledge.

Dion, Douglas. 1997. Evidence and Inference in the Comparative Case Study. *Comparative Politics* 30 (2):127-145.

Ebbinghaus, Bernhard. 2005. When Less is More: Selection Problems in Large-N and Small-N Cross-national Comparisons. *International Sociology*: forthcoming.

Fullbrook, Edward, ed. 2003. *The Crisis in Economics: The Post-autistic Economics Movement*. London: Routledge.

Geddes, Barbara. 2003. *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*. Ann Arbor: University of Michigan Press.

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

Gerring, John. 2004. What is a Case Study and What is it Good for? *American Political Science Review* 98 (2):341-354.

Gillies, Donald. 2000. *Philosophical Theories of Probability*. London: Routledge.

Goertz, Gary, and Harvey Starr. 2003. *Necessary Conditions: Theory, Methodology, and Applications*. Boulder: Rowman & Littlefield.

Goldthorpe, John H. 2000a. Causation, Statistics, and Sociology. In *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*, edited by J. H. Goldthorpe. Oxford: Oxford University Press.

Goldthorpe, John H. 2000b. Current Issues in Comparative Macrosociology. In *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*, edited by J. H. Goldthorpe. Oxford: Oxford University Press.

Granato, Jim, and Frank Scioli. 2004. Puzzles, Proverbs, and Omega Matrices: The Scientific and Social Significance of Empirical Implications of Theoretical Models (EITM). *Perspectives on Politics* 2 (2):313-323.

Hall, Peter A. 2003. Aligning Ontology and Methodology in Comparative Research. In *Comparative Historical Research in the Social Sciences*, edited by J. Mahoney and D. Rueschemeyer. Cambridge: Cambridge University Press.

Hofstadter, Douglas R. 1979. *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books.

Huber, Evelyne, and John Stephens. 2001. *Development and Crisis of the Welfare State*. Chicago: The University of Chicago Press.

King, Gary, Robert Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
Levi, Margaret. 1997. A Model, a Method, and a Map: Rational Choice in Comparative and Historical Analysis. In *Comparative Politics. Rationality, Culture, and Structure*, edited by M. I. Lichbach and A. S. Zuckerman. Cambridge: Cambridge University Press.
Lieberson, Stanley. 1997. The Big Broad Issues in Society and Social History: Application of a Probabilistic Perspective. In *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*, edited by V. R. McKim and S. P. Turner. Notre Dame: University of Notre Dame Press.
Lijphart, Arend. 1997. Reflections on the First Twenty-Five Years of the European Journal of Political Research. *European Journal of Political Research* 31 (1-2):5-16.
Little, Daniel. 2000. Explaining Large-Scale Historical Change. *Philosophy of the Social Sciences* 30 (1):89-112.
Mahoney, James. 2004. Comparative-Historical Methodology. *Annual Review of Sociology* 30:81-101.
Mahoney, James, and Dietrich Rueschemeyer. 2003. Comparative Historical Analysis: Achievements and Agendas. In *Comparative Historical Analysis in the Social Sciences*, edited by J. Mahoney and D. Rueschemeyer. Cambridge: Cambridge University Press.
Mäki, Uskali, ed. 2002. *Fact and Fiction in Economics: Models, Realism, and Social Construction*. Cambridge: Cambridge University Press.
Marsh, David, and Heather Savigny. 2004. Political Science as a Broad Church: The Search for a Pluralist Discipline. *Politics* 24 (3):155-168.
Mayntz, Renate. 2002. Zur Theoriefähigkeit makro-sozialer Analysen. In *Akteure, Mechanismen, Modelle: Zur Theoriefähigkeit makro-sozialer Analysen,* edited by R. Mayntz. Frankfurt/M.: Campus.
Moses, Jonathon W., Benoît Rihoux, and Bernhard Kittel. 2005. Mapping Political Methodology: Reflections on a European Perspective. *European Political Science* 4 (1):1-14.
Norris, Pippa. 1997. Towards a More Cosmopolitan Political Science. *European Journal of Political Research* 31 (1-2):17-34.
Pierson, Paul. 1994. *Dismantling the Welfare State? Reagan, Thatcher and the Politics of Retrenchment.* Cambridge: Cambridge University Press.
Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley.
Ragin, Charles C. 1992. Introduction: Cases of "What is a Case?" In *What is a case?: Exploring the foundations of social inquiry*, edited by C. C. Ragin and H. S. Becker. Cambridge: Cambridge University Press.
Ragin, Charles C. 2000. *Fuzzy-set Social Science*. Chicago: University of Chicago Press.
Rosenberg, Alexander. 1992. *Economics: Mathematical Politics or Science of Diminishing Returns*. Chicago: University of Chicago Press.
Salmon, Wesley C. 1998. *Causality and Explanation*. Oxford: Oxford University Press.
Sayer, Andrew. 1984. *Method in Social Science. A Realist Approach*. London: Hutchison.
Scharpf, Fritz W. 2000. Institutions in Comparative Policy Research. *Comparative Political Studies* 33 (6-7):762-790.
Skocpol, Theda, and Margaret Somers. 1980. The Uses of Comparative History in Macrosocial Inquiry. *Comparative Studies in Society and History* 22:174-197.
Smelser, Neil. 1976. *Comparative Methods in the Social Sciences*. Englewood Cliffs: Prentice Hall.
Tilly, Charles. 1984. *Big Structures, Large Processes, Huge Comparisons*. New York: Russell Sage Foundation.
Waldner, David. 2002. "Anti Anti-Determinism: Or What Happens When Schrodinger's Cat and Lorenz's Butterfly Meet Laplace's Demon in the Study of Political and Economic Development." Paper presented to the annual meetings of the American Political Science Association, Boston MA (August-September).
Watkins, J. W. N. 1952. Ideal Types and Historical Explanation. *British Journal of the Philosophy of Science* 3 (9):22-43.
Weber, Max. 1949. 'Objectivity' in Social Science and Social Policy. In *The Methodology of the Social Sciences*, edited by M. Weber. New York: The Free Press.
Weber, Max. 1978. *Economy and Society. An Outline of Interpretive Sociology*. Berkeley: University of California Press.
Yanow, Dvora. 2003. Interpretive Political Science: What Makes This Not a Subfield of Qualitative Methods. *Qualitative Methods* 1 (2):9-13.
Zukav, Gary. 1979. *The Dancing Wu Li Masters. An Overview of the New Physics*. New York: Bantam.

Clarifying Comparative-Historical Methodology

James Mahoney
Brown University
James_Mahoney@Brown.edu

Bernard Kittel discusses "three contentious issues which seem to be regarded as uncontroversial in the American debate." These issues are: (a) knowledge accumulation, (b) generalization, and (c) probabilistic causation. Kittel argues that qualitative analysts–and comparative-historical researchers in particular–"unnecessarily weaken their case. . . by accepting the quantitative template in these respects." His overall argument is presented in a provocative manner; it is an effort to move the discussion forward in the ongoing debate about methodology in the United States.

In this reply, I agree with many of Kittel's reservations about the mainstream statistical template, but I argue that he mischaracterizes the degree to which comparative-historical analysis actually follows this template. In addition, I express reservations about certain conclusions and directions that Kittel offers for future macro-comparative work.

Knowledge Accumulation

Kittel has "no objections to the general goal of cumulation," but he is concerned that researchers explicitly seeking cumulation may actually get in the way of intellectual progress. In his view, the explicit quest for knowledge accumulation may, ironically, inhibit the actual achievement of knowledge generation. To support this claim, he looks to the natural sciences, where intellectual progress often has occurred in the absence of any single theoretical paradigm and has been stimulated precisely by the ability of scholars to break out of reigning theoretical straightjackets.

I believe that there is much to be said for Kittel's view that it is inappropriate for scholars to try to present a single theoretical approach or methodological template as the sole means to achieve knowledge gains in political science. In the introduction to *Comparative Historical Analysis in the Social Sciences*, for example, Dietrich Rueschemeyer and I expressed reservations about rational choice theory for pre cisely this reason. Likewise, we argued that "comparative historical analysts are decidedly pluralist in their use of overarching theories . . . they do not hesitate to seek guidance from a range of theoretical traditions" (Mahoney and Rueschemeyer 2003: 21-22). In our view, comparative-historical researchers must always feel free to draw on any methodological or theoretical tools that will enable them to validly answer the questions at hand. Indeed, qualitative methodology more generally is hardly converging on any single approach or an exclusive set of tools as the sole means to social science knowledge; rather, the emphasis is very much on the diversity of tools that should and do inform research (e.g., Brady and Collier 2004; George and Bennett 2005; Pierson 2004).

When compared to Kittel, I am more convinced that cumulation has been important to knowledge generation in the social sciences, and in comparative-historical analysis in particular. I think this difference in opinion is linked in part to our contrasting understandings of knowledge accumuluation. As I define it, knowledge accumulation refers to a process in which new knowledge claims (e.g., descriptive findings and causal findings) and new knowledge generation tools (e.g., methodologies) grow out of and depend on previously existing knowledge and tools (see Mahoney 2003 for a discussion). Knowledge accumulation does not assume that an analyst works within a single theoretical or methodological program. In fact, knowledge accumulation requires a scholar to show creativity in combining preexisting work with new techniques, new frameworks, and/or new ideas.

In the field of comparative historical analysis, substantive knowledge accumulation has occurred across diverse areas, including studies of revolutions, welfare states, and democracy and authoritarianism (see Goldstone 2003; Amenta 2003; Mahoney 2003). Initial studies of these topics spurred new research, and this new research led to the creation of new findings and tools. Along the way, many puzzles were solved even as new research areas were discovered. As a result, we now know much more about revolutions, welfare states, democracy and authoritarianism, and several other research topics than we did twenty or thirty years ago.

More generally, cumulative research programs encourage scholars to use and learn from the insights of others rather than to start from scratch in their research. As I put it: "To the degree that one believes that scholars can learn from one another and that learning facilitates knowledge generation, cumulative research communities have an inherent advantage over their noncumulative rivals" (Mahoney 2003:164). Thus, whereas both Kittel and I agree that researchers must not limit themselves to some narrow set of theoretical guidelines or methods in the pursuit of knowledge accumulation, I am more inclined to believe that scholars have achieved knowledge accumulation by creatively drawing on and elaborating the insights of previous scholars.

Generalizability

Kittel argues that "recent work in the field of comparative-historical analysis has become obsessed with theory testing and generalizability." This obsession, in turn, has led to three key problems. First, Kittel suggests that comparative-histori-cal analysts have sought "causal laws" by looking at the origins of singular events in macro units like states. The problem is that the objects of study in this research are not suitable to the development of any such causal laws or general regularities; rather, singular outcomes should be explained through other techniques not designed to account for regularities.

In my view, Kittel presents too stark of a contrast between research that focuses on a single case and research that seeks to generalize to large populations. Comparative-historical analysis stands between these two alternatives; it is a mode of research in which scholars compare multiple cases in the effort to formulate causal generalizations about those cases. The causal generalizations that emerge are not "causal laws" in the sense of ahistorical generalizations that apply across all times and places. Rather, the generalizations are bounded by particular scope conditions that define comparable cases. With these bounded generalizations, comparative-historical researchers solve certain kinds of focused puzzles and contribute to substantive knowledge in ways that single-case study analyses and large-N analyses may not.

Second, Kittel argues that comparative-historical researchers tend to understand cases as theoretical constructs (e.g., "I have a case of revolution") but introduce inconsistency when they treat cases as particular spatial-temporal units (e.g., "My case is France"). In fact, however, comparative-historical researchers may define cases in much the same way as statistical researchers–i.e., cases are units of analysis (e.g., states, regions) across which variables are measured. A possible difference with much statistical research is that comparative-historical researchers try to carefully examine the assumption that all units under investigation are the same basic type of unit, such that they can be meaningfully compared with one another. Defining comparable populations of cases (given a particular theory) therefore becomes a crucial part of comparative-historical research (see Ragin 2000).

Third, Kittel argues that the cases in comparative-historical research interact with and affect one another so much that one cannot plausibly treat them as independent and homogeneous units. The issue here is sometimes called "Galton's problem," and Kittel is not the first European scholar to raise it in relationship to comparative-historical analysis (see Goldthorpe 1997). The problem in fact applies to all macro research, but as Rueschemeyer and Stephens (1997) note, comparative-historical analysis offers especially useful tools for dealing with it: "The reason is that historical research can trace a case over time and take full account of the way in which the characteristics that may or may not have been the result of diffusion are linked to their local context" (Rueschemeyer and Stephens 1997: 61). Thus, comparative-historical researchers do not have to pretend that diffusion does not take place; rather, they can actively build it into their theories and examine it closely through a careful processual analysis of individual cases.

Determinism and Probabilism

Kittel usefully builds on George and Bennett's (2005) distinction between inherent probabilism and complexity-induced probabilism. And Kittel may be right that many comparative-historical researchers assume that, ontologically, the macro world is deterministic and that unexplained outcomes or variance results from deficits in theory. Indeed, some comparative-historical theories are deterministic—that is, the analyst presents a theory that seeks to identify one or more combinations of variable values that are sufficient for the outcomes to be explained (within the comparable population of cases under investigation). An outcome in a single case that is not adequately explained may be considered a real problem for these comparative-historical researchers.

Yet other comparative-historical researchers assume complexity-induced probabilism. Kittel's observations raise two key issues in this regard. The first concerns the assumption that one needs a large N to test probabilistic hypotheses. Although this assumption is true in quantitative research, it is not true for many kinds of comparative-historical research. For example, so long as one has a medium number of cases (e.g., 15 or 20 cases), it is certainly possible to draw conclusions such as "X is necessary for Y 80 percent of the time" that pass standard statistical significance tests (e.g., Dion 1998, Ragin 2000, Braullmoeler and Goertz 2000). This point is poorly understood, so it is worth emphasizing it: one does not need a large sample, only a small to medium N, as is sometimes used in comparative-historical research, to draw statistically significant conclusions about probabilistic necessary and sufficient causation. Still, if one is working with a very small N (e.g., N = 2 or 3), then one would have to draw probabilistic inferences by relying primarily on within-case analysis rather than cross-case comparison.

The more challenging issue concerns the use of probabilistic causation when the sample of cases is equal to the full population of cases about which one wishes to generalize. What does it mean to say that some variable is a probabilistic cause when there is no larger population about which one is generalizing? For Kittel and at least some statisticians, it is not meaningful to speak of probabilities when one is not generalizing from a random sample to a population. My own view is that the tendency of some comparative-historical researchers to seek deterministic explanations that achieve causal sufficiency is closely related to the fact that they may not be generalizing about a larger population. By contrast, when comparative-historical researchers seek to develop arguments that can be readily generalized to a larger population of cases, they are more likely to shift to a probabilistic mode of inference.

Should We Build from the Micro Level?

Kittel concludes his essay by suggesting that scholars who study macrophenonmena should build explanations by piecing together regularities at the micro level. As he puts it: "We should build an explanatory story by referring to a sequence of regularity statements at the micro level which are connected by narrative elements." As he suggests, this research agenda is congruent with the emphases of many scholars who advocate analytic narratives (e.g., Bates et al. 1998) and partly overlaps with those who view the identification of "mechanisms" as involving a focus on the micro level.

Although a concern with micro foundations is valuable, I believe that the overall project of "building up" from the micro level to coherent and overall explanations of macrophenomenon will not succeed. Rather, the best works we have in comparative-historical analysis are studies that identify macro-level causal regularities. And these works do not start with individual regularities to identify the macro patterns. Instead, the macro patterns typically are developed through the use of theory and historical evidence at the macro level itself. For example, Skocpol (1979) did not develop her hypotheses about the effects of international pressures, state autonomy, and peasant community solidarity on social revolution by beginning with regularities at the micro level. Rather, she employed theory and evidence at the macro level itself.

To be sure, once macro level hypotheses are formulated, it is often useful to "build down" to more micro-level processes and thereby test whether hypotheses still make sense in light of detailed evidence from within cases. Yet, this use of micro-level narrative is subservient to and fundamentally structured by the macro-level analysis. Thus, for instance, Skocpol's (1979) narrative does certainly make reference to regularities in individual behavior (e.g., monarchs will respond to crises by attempting modernizing reforms; landlords will respond to crises by squeezing peasants; peasants will respond to crises with revolts if they have the capacity to do so). But the macro-level analysis directed her attention to these micro-level regularities in the first place.

Kittel would disagree with my skepticism regarding work that starts at the micro level and tries to build toward a compelling overall explanation of macrophenomena. In the final analysis, though, the proof is in the pudding. At this point, scholars who start with macro-level regularities and then supplement these insights with a concern with micro foundations can point to major achievements. I do not believe the same can yet be said for those who try to build exclusively from the micro level in developing their explanations of macro phenomenon.

References

Amenta, Edwin. 2003. "What We Know about the Development of Social Policy: Comparative and Historical Research in Comparative and Historical Perspective." Pp. 91-130 in *Comparative Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press.

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

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

Braumoeller, Bear F., and Gary Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science* 44: 844-858.

Dion, Douglas. 1998. "Evidence and Inference in the Comparative Case Study." *Comparative Politics* 30:127-146.

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

Goldstone, Jack A. 2003. "Comparative Historical Analysis and the Knowledge Accumulation in the Study of Revolutions." Pp. 41-90 in *Comparative Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press.

Goldthorpe, John H. 1997. "Current Issues in Comparative Macrosociology: A Debate on Methodological Issues." *Comparative Social Research* 16:1-26.

Mahoney, James. 2003. "Knowledge Accumulation in Comparative Historical Research: The Case of Democracy and Authoritarianism." Pp. 131-174 in *Comparative Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press.

Mahoney, James, and Dietrich Rueschmeyer. 2003. "Comparative Historical Analysis: Achievements and Agendas." Pp. 3-38 in *Comparative Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press.

Pierson, Paul. 2004. *Politics in Time*. Princeton: Princeton University Press.

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

Rueschemeyer, Dietrich and John D. Stephens. 1997. "Comparing Historical Sequences–A Powerful Tool for Causal Analysis." *Comparative Social Research* 17:55-72.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.

Symposium II: Necessary Conditions

Necessary Condition Hypotheses as Deterministic or Probabilistic: Does it Matter?[1]

Gary Goertz
University of Arizona
ggoertz@u.arizona.edu

That necessary condition hypotheses are deterministic forms part of the methodological and theoretical folk wisdom of the social sciences. When mentioned, it usually serves as means for a quick dismissal of the hypothesis in question, or as a segue to other issues. This essay proposes that the folk theorem "necessary conditions are deterministic" avoids the central issues which are causal mechanisms, functional form, and testing philosophy. These are the core topics since the two corollaries of necessary condition folk theorem are that (1) one counter-example suffices to reject a necessary condition hypothesis and (2) the world is probabilistic and hence deterministic theories are not useful. The first claim is about empirical testing philosophy, what one could call a naive falsificationist view.[2] The second deals with the nature of the world, what Hall (2003) calls an ontological position about what kinds of theories will be most useful to (social) scientists. I shall argue that the folk theorem about necessary condition hypotheses and its two corollaries do not survive serious analysis.

The character of necessary condition hypotheses merits debates since they arise with great regularity in social science (see Goertz 2003 for a list of 150 of them, representing a wide range of substantive areas and methodological orientations). It is all the more important for qualitative theory and methodology since these theorists appear to have an elective affinity for this kind of thinking and causal explanation (see Mahoney 2004 for a discussion in the context of comparative-historical methods and theory). Many of the important works of comparative politics and sociology can be cited in this regard: classics such as Radcliffe-Brown (1952), Lipset (1959), Malinowski (1944), Gerschenkron (1962), Moore (1966), and Skocpol (1979), and more recent work in the same tradition such as Downing (1992), Goodwin (2001), Jacoby (2000), Linz and Stepan (1996), Waldner (1999), and Weldon (2002).

Typical of qualitative-comparative studies is a desire to explain individual cases. While good scholars have always been interested in theories that work in multiple settings, there is a constant concern among qualitative theorists that the theory work well for central cases. If a theory of revolution does not work well for the French or Russian Revolution then it will not be well-received even if it does very well on other, historically prominent cases. To use Ragin's terminology, qualitative scholars are more case-oriented than variable-oriented quantitative scholars, who are generally unconcerned about how well their theories work for any given observation. There is no doubt that causal explanations in individual cases often apply the basic necessary condition counterfactual "if X had not occurred or been present then Y would not have occurred" (see Goertz and Levy 2004 for extensive examples from the literatures on the causes of World War I and the end of the Cold War). In short, historical counter-factuals often take the necessary condition form.

One cannot dismiss the affinity of qualitative scholars for necessary conditions on the basis of lack of rigorous thinking. Prominent necessary condition hypotheses occur (though perhaps less frequently) in formal and game-theoretic work; for example, they occupy the core of Bueno de Mesquita's most influential work (1981; Bueno de Mesquita and Lalman 1992). Other examples include Riker and Sened (1991), Pahre (1997), Krehbiel (1985), Niou, Ordeshook, and Rose (1989), and Taylor (1987). Formal theory almost by its very nature produces theorems containing necessary conditions.

The Appendix to this essay provides a small sample of prominent necessary condition hypotheses. As it illustrates, they arise across the complete range of N, large, small, and case study. They occur in theoretical perspectives ranging from game theory to social constructivism. However, there does appear to exist a particular affinity between qualitative scholars and necessary condition hypotheses. Roughly, as the N of the study increases, the more likely it is that the author will express hypotheses in probabilistic language. Conversely, in small-N or case studies one is much more likely to find causal explanations using necessary and sufficient condition language. In short, necessary condition hypotheses, their theoretical status, and how they are to be evaluated empirically deserve attention, particularly within the context of qualitative methods.

What is the Status and Usefulness of Deterministic Theories?

It is useful to situate the specific claim about necessary conditions as deterministic within the larger context of deterministic theories in general. We can provisionally define as deterministic all theories that do not have an explicit probabilistic or stochastic component (this need not necessarily be an error term).

Lieberson (1991; 2004; see also Goldthorpe 1997, 6) provides an example of someone for whom the fact that the methodology/theory is deterministic serves as grounds for criticism and rejection. He has repeatedly argued that Ragin's methodologies–Boolean (1987) and fuzzy-set (2000)–are not useful because they are deterministic. Lieberson is thus applying what I have called the ontological corollary to the deterministic folk theorem. He sees the world as probabilistic in nature, hence deterministic methodologies have lesser, if any, value.

If, however, one surveys the natural and social sciences, there is a large range of nonprobabilistic theories that have been extremely useful and/or influential over the centuries and recent decades. Natural sciences had a history of success, e.g., Newtonian physics, long before probability theory was even invented. Calculus and differential equations have been widely applied in the natural sciences and have a long tradition in economics. Within international relations one cannot forget the groundbreaking work of Richardson on arms races (1960; for other and recent applications of dynamic equation models, see Kadera 2001). Much of game theory is deterministic. It seems hard to deny the usefulness of deterministic models in general. It can be claimed that set theory and logic (i.e., the mathematics of necessary conditions) are not useful for explaining or modeling social phenomena, but it cannot be on the grounds of determinism.

We can think of different kinds of mathematics as representing different kinds of causal mechanisms or explanations. More concretely, we can present this as a debate about functional form. Suppose we take two different deterministic models:

$$Y = B_0 + B_1X_1 \qquad (1)$$

$$X \text{ is necessary for } Y \qquad (2)$$

For a given phenomenon we can ask which is a better model or explanation. We can ask if the algebra and calculus of equation (1) works better than the logic and set theory of equation (2). Clearly, the linear (and additive, if more variables are included) model of equation (1) is familiar to students of statistics, while the mathematics of equation (2) are rarely taught.[3] Presented in this manner, the fact that both equations are deterministic hardly seems relevant to assessing the value of competing models.

Evaluating Deterministic Models

A key corollary of the deterministic folk theorem involves the claim that one counter-example suffices to reject a necessary condition hypothesis. We can ask two questions in this regard: First, is that a reasonable testing philosophy? and second, is that how deterministic theories are usually tested? But this avoids the most important question: which explanation, equation (1) or equation (2), fits the data better? Since equations (1) and (2) represent different functional forms (i.e., different causal mechanisms), we can skip philosophy and move to the pragmatic question of comparing theory to data. Figures 1 and 2 illustrate situations where the data fit each theory well. If the linear causal mechanism generates the data then we would see the data scattered as in Figure 1; if the causal mechanism involves necessary conditions, then we should see a scatter plot like that in Figure 2 (see Mahoney 2004 for a similar discussion). The differences between these scatter plots reflect the fact that we have different causal explanations and different functional forms.

In terms of fit, the data in Figure 2 fit the necessary condition hypothesis of equation (2) much better than the data of Figure 1 fit the linear model of equation (1). The deterministic linear model requires all the data points to lie on the line. In contrast, the necessary condition hypothesis as modeled by fuzzy sets requires that all points lie on or below the X = Y diagonal (see Ragin 2000 for an extensive discussion of this). While most of the points lie near the line not that many lie on it in Figure 1, while in Figure 2 only three points (indicated by "?") lie above the diagonal.[4]

One would not reject the necessary condition hypothesis just because there are three points that are above the diagonal, just as one would find that the data in Figure 1 quite strongly support the linear model even though most of the points are not on the line ($R^2 = .85$) This is true for the real-life testing of all deterministic theories. One does not expect them to fit perfectly; researchers are satisfied if the data are "close enough" to what the theory expects. In summary, the one counterexample and reject corollary is never applied to deterministic theories, and there is no reason to apply it to necessary condition ones.

Figure 1: Linear Functional Form

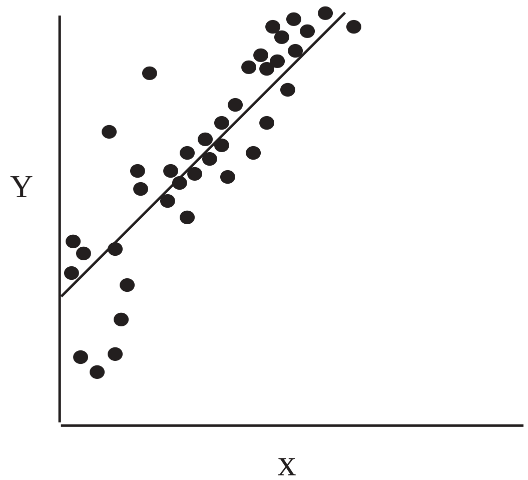
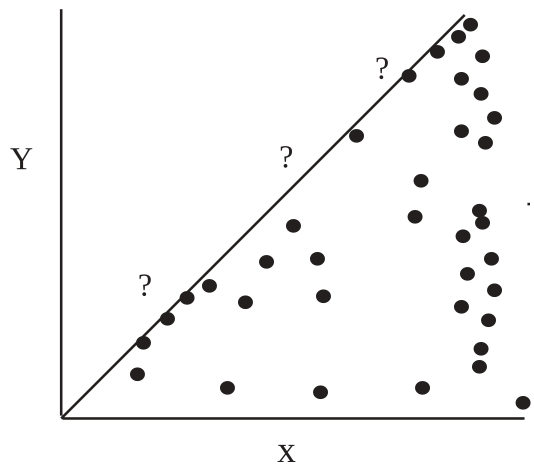


Figure 2: Necessary Condition Functional Form (from Fuzzy Sets)



Are Necessary Condition Hypotheses Deterministic?

The previous sections have addressed the corollaries of the folk theorem that necessary condition hypotheses are deterministic, but I have yet to address the theorem itself. One might say that by my own definition of deterministic theories–i.e., no explicit probabilistic components–it is clear that the answer to the question must be yes. I would like to suggest in this section that things are not so clear as they might seem.

I have stressed throughout this short essay that what we should be focusing on is causal explanations, mechanisms, and functional form. If we can express the basic causal claim of a necessary condition in statistical or probabilistic language, then the whole debate would tend to become relatively moot. One can take the analogy of algebra and geometry. Many theorems of algebra can be expressed geometrically and vice versa. Is the same true of necessary condition hypotheses?

Expressing necessary condition hypotheses in probabilistic terms is the tack taken by Braumoeller and Goertz (2000). As far as I can tell no one contests the following as either a definition of a dichotomous necessary condition hypothesis or as a direct implication of it:

$$P(X = 1|Y = 1) = 1 \tag{3}$$

If X is necessary for Y, then whenever Y occurs we must find that X preceded it.[5] One might object that I am playing with words here since determinism can be phrased as "probability equals 1." The question I would like to raise here is "does the nature of causal explanation change when the probability is .99?" A variety of authors suggest that there is not a dramatic difference. Little says: "If C is a necessary condition for E, then the probability of E in the absence of C is zero (P(E|¬C) = 0). If C is a sufficient condition for E, then the probability of E in the presence of C is one (P(E|C) = 1). And we can introduce parallel concepts that are the statistical analogues of necessary and sufficient conditions. C is an *enhancing* causal factor just in case P(E|C) > P(E), and C is an *inhibiting* causal factor just in case P(E|C) < P(E). The extreme case of an inhibiting factor is the absence of a necessary condition, and the extreme case of an enhancing causal factor is a sufficient condition" (Little 1991, 27; for a similar treatment see Owens 1992: 5).[6]

Sometimes, however, the move from a deterministic to a probabilistic model can have major consequences. One characteristic of game-theoretic models is that situations of uncertainty often have quite different strategic properties than do situations of certainty. For example, Morton (1999) discusses spatial voting models where there are significant differences between deterministic models and probabilistic ones. As a result, the whole theory changes, and we would not say that the case where the probability is 1.00 is basically the same as when the probability is .99. At the same time, I think there are many situations where the basic hypothesis does really not change much when one moves from "with certainty" to "is extremely likely."

One way to see if the move from 1.00 to .99 matters is to examine how testing of deterministic theories is conducted for point hypotheses similar to equation (3). One large body of literature comes from experimental economic studies of bargaining behavior (e.g., Roth 1995 for a survey; see Morton 1999 for a lengthy discussion of point predictions in formal models). Frequently the expected utility model (deterministic) predicts a unique equilibrium behavior. Often in the context of prisoners' dilemma–like situations it is no cooperation by anyone. One of the notable findings of this literature is how often people cooperate when the theory says they should not (often up to 50 percent do). Economists work hard to get this percentage down, and consider themselves successful if cooperation rates are less than 20 percent.[7] Certainly the experimental economists do not make a big difference between a couple of percentage points: for them .99 is not significantly different from 1.00.

Another possibility is to express the basic ideas of Boolean logic (e.g., logical ANDs and ORs) in probabilistic terms. Cioffi-Revilla and Starr (2003; see also Cioffi-Revilla 1998; see Braumoeller 2003 for some statistics) model complex necessary and sufficient condition theories in probabilistic terms. They are extremely clear about the use of Boolean logic and hence necessary and/or sufficient conditions. For example,

AXIOM 1 (First-Order Causality: Political Necessity). *A political behavior event B in U occurs when W and O occur. Formally, B is defined by the causal equation*

$$B = W \text{ AND } O \tag{4}$$

where the Boolean AND connective stands for the formal logic conjunction (i.e., "W AND O are necessary conditions for B").

THEOREM 1 (First-Order Probability of Political Behavior). *The first-order probability of political behavior B(W,O) is equal to the product of the probability of willingness W and the probability of opportunity O. Formally,*

$$\Pr(B) = \Pr(W \cap O) = \Pr(W) \times \Pr(O) \tag{5}$$
$$B \equiv WO \tag{6}$$
$$= P^2 \tag{7}$$

in the special case where P = W = O.

It is extremely useful to have the basic claims of a necessary condition hypothesis expressed in probabilistic or statistical terms. This brings out even more clearly the fundamental differences in functional form. The Braumoeller and Goertz (2000) approach also illustrates the differences between a necessary condition hypothesis and typical statistical ones once they are speaking the same language. The hypothesis expressed in equation (3) is a point prediction, $p = 1.00$, or a narrow range prediction, i.e., $p \sim 1$. Contrast this with the typical null hypothesis model where the alternative hypothesis (what the researcher wants to prove) is typically $p \neq .5$, or $p > .5$.

In expressing necessary condition hypotheses in different forms, I have shown that they have a variety of testable implications (e.g., point hypotheses, triangular scatterplots, counterfactual hypotheses). This should be the focus of the methodological discussion. The folk theorem and its two corollaries about the determinism of necessary condition hypotheses crumble under analysis. Many have applied standards of theory and testing to necessary condition hypotheses that would never be applied to other deterministic hypotheses. At the same time, it is clear that the basic claims of necessary condition hypotheses can be expressed in probabilistic or statistical terms. The title of this essay posed a question; the answer is "no, it usually does not matter."

Appendix

1. "The expected utility model provides a framework from which at least several significant, several lesser, deductions about the necessary conditions for war have been made. . . . (Bueno de Mesquita 1981: 92) The results just reported strongly support the proposition that positive expected utility is necessary–though not sufficient–for a leader to initiate a serious international dispute, including a war." (ibid: 129)

2. "The concept of historical prerequisites of modern industrialization is a rather curious one. Certain major obstacles to industrialization *must* be removed and certain things propitious to it *must* be created before industrialization can begin. . . Along with it goes the idea of the uniformity of industrial development in the sense that every industrialization necessarily must be based on the same set of preconditions" (Gerschenkron 1962: 31; note that the title of the essay is "Reflections on the concept of 'prerequisites' of modern industrialization").

3. "*Theorem 4.1: Communication leads to enlightenment if and only if:* 1. the speaker is persuasive, 2. only the speaker initially possesses the knowledge that the principal needs, and 3. neither common interests nor external forces induce the speaker to reveal what he knows" (Lupia and McCubbins 1998: 69).

4. Kiser and Levi (1996: 189–90) describing Moore, Skocpol, and Goldstone:

> Moore: Only where there was a relatively strong bourgeoisie independent of the state and only where the aristocracy and peasantry either sided with the bourgeoisie or were negligible was there a revolution that led to democracy.
> Skocpol: Only where there are pressures on states due to wars and international competition and only where these pressures result in a conjuncture of fiscal crisis, abandonnment of the state by the dominant classes, and peasant revolts based in strong peasant communities is there the possibility of a social revolution.
> Goldstone: Only where there have been demographic shifts and increasing demographic pressures that create political stress (elite competition, fiscal crisis, and mass mobilization potential based on concentration of youth) will there be state breakdown. When the cultural framework permits the development of an elite ideology committed to innovation there will be a revolution.

5. "To recapitulate, the Sarajevo assassinations changed the political and psychological environment in Vienna and Berlin in six important ways, all of which were probably necessary for the decisions that led to war" (Lebow 2000: 605).

6. "Clearly, the necessary but not sufficient conditions for major war emerge only in the rare instances when power parity is accompanied by a challenger overtaking a dominant nation. The odds of a war in this very reduced subset are 50 percent. No other theoretical statement has, to our knowledge, reduced the number of cases to such a small set, and no other is so parsimonious in its explanatory requirements." (Organski and Kugler 1989: 179).

7. "By 'design principle' I mean an essential element or condition that helps to account for the success of these institutions [common pool resource] in sustaining the common pool resources and gaining the compliance of generation after generation of appropriators to the rules in use. . . I am willing to speculate. . . [that] it will be possible to identify a set of

necessary design principles and that such a set will contain the core of what has been identified here" (Ostrom 1991: 90–91).

8. "In this study, I shall argue that people go out into the streets and protest in response to deeply felt grievances and opportunities. But this produces a protest cycle only when structure cleavages are both deep and visible and when opportunities for mass protest are opened up by the political system" (Tarrow 1989: 13).

9. "Expressed as a necessary condition, my argument is that the absence of high levels of elite conflict is necessary for sustained economic development and industrial transformation" (Waldner (1999: 16; see also Figure 1).

10. "Decline was a necessary condition of change [in Soviet policy], but clearly insufficient to determine the precise nature of change. The fact that change in relative power can lead to more than one behavioral response does not mean that it is not an important part of the explanation" (Wohlforth, 1995: 186; Wohlforth later argues that it is better to express this in probabilistic terms, see Brooks and Wohlforth 2002).

Notes

[1] I would like to thank Bear Braumoeller, Lars-Erik Cederman, Claudio Cioffi-Revilla, Jim Mahoney, and Charles Ragin for comments on an earlier draft of this paper.

[2] For example, "When we say 'If X_1 then Y ,' we are making a deterministic statement. When we say, 'The presence of X_1 increases the likelihood or frequency of Y,' we are making a probabilistic statement. Obviously, if given the choice, deterministic statements are more appealing. They are cleaner, simpler, and more easily disproved than probabilistic ones. One negative case (Y's absence in the presence of X_1) would quickly eliminate a deterministic statement" (Lieberson 1997, 364; see also 2004). Alker makes the same claim: "The easily falsifiable nature of necessary and/or sufficiency arguments – all that is required is a counter-example – is surely one of the most at-tractive features of requisites analysis from an empirical point of view" (Alker 1970, 879).

[3] In philosophy of course it is the opposite, everyone learns logic but only a small subset learn probability and statistics.

[4] It is thus not surprising that when two different functional forms are applied to the same data that one can get different results (e.g., Koenig-Archibugi 2004).

[5] Among other things, it is this fact that Dion (2003) utilizes to argue that selecting on the dependent variable is legitimate for necessary conditions. The typical statistical model is of the form E(Y|X), in contrast the necessary cause model conditions on Y not X.

[6] Elster sees scientific laws has having a sufficient condition form: "a [covering, scientific] law has the form "IF conditions C_1,C_2, . . . ,C_n obtain, THEN always E" (1999, 5). As is typical, he is more interested in the "always" aspect of this definition of covering laws than the sufficient condition functional form.

[7] In comparison the Braumoeller and Goertz (2000) or Ragin (2000) probabilistic standards are much more stringent.

References

Alker, H. 1970. Integration logics: a review, extension and critique. *International Organization* 24:869–913.

Braumoeller, B. 2003. Causal complexity and the study of politics. *Political Analysis* 11:209–33.

Braumoeller, B., and G. Goertz. 2000. The methodology of necessary conditions. *American Journal of Political Science* 44:844–58.

Brooks, S., and W. Wohlforth. 2002. From old thinking to new thinking in qualitative research. *International Security* 26:93–111.

Bueno de Mesquita, B. 1981. *The war trap.* New Haven: Yale University Press.

Bueno de Mesquita, B., and D. Lalman. 1992. *War and reason: domestic and international imperatives.* New Haven: Yale University Press.

Cioffi-Revilla, C. 1998. *Politics and uncertainty: theory, models and applications.* Cambridge: Cambridge University Press.

Cioffi-Revilla, C., and H. Starr. 2003. Opportunity, willingness, and political uncertainty: theoretical foundations of politics. In G. Goertz and H. Starr (eds.) *Necessary conditions: theory, methodology, and applications.* New York: Rowman & Littlefield.

Dion, D. 2003. Evidence and inference in the comparative case study. In G. Goertz and H. Starr (eds.) *Necessary conditions: theory, methodology, and applications.* New York: Rowman & Littlefield.

Downing, B. 1992. *The military revolution and political change: origins of democracy and autocracy in early modern Europe.* Princeton: Princeton University Press.

Elster, J. 1999. *Strong feelings: emotion, addiction, and human behavior.* Cambridge: MIT Press.

Gerschenkron, A. 1962. Reflections on the concept of 'prerequisites' of modern industrialization. *Economic backwardness in historical perspective: a book of essays.* Cambridge.: Cambridge University Press.

Goertz, G. 2003. The substantive importance of necessary condition hypotheses. In G. Goertz and H. Starr (eds.) *Necessary conditions: theory, methodology, and applications.* New York: Rowman & Littlefield.

Goertz, G., and J. Levy (eds.). under review. *Causal explanations, necessary conditions, and case studies: World War I and the end of the Cold War.* London: Palgrave.

Goldthorpe, J. 1997. Current issues in comparative macro-sociology: a debate on methodological issues. *Comparative Social Research* 16:1–26.

Goodwin, J. 2001. *No other way out: states and revolutionary movements, 1945–1991.* Cambridge: Cambridge University Press.

Hall, P. 2003. Aligning ontology and methodology in comparative research. In J. Mahoney and D. Rueschemeyer (eds.) *Comparative historical analysis in the social sciences.* Cambridge: Cambridge University Press.

Jacoby, W. 2000. *Imitation and politics: redesigning modern Germany.* Ithaca: Cornell University Press.

Kadera, K. 2001. *The power-conflict story: a dynamic model of interstate rivalry.* Ann Arbor: University of Michigan Press.

Kiser, E., and M. Levi. 1996. Using counterfactuals in historical analysis: theories of revolution. In P. Tetlock and A. Belkin (eds.) *Counterfactual thought experiments in world politics: logical, methodological, and psychological perspectives.* Princeton: Princeton University Press.

Koenig-Archibugi, M. 2004. Explaining government preferences for institutional change in EU foreign and security policy. *International Organization* 58:137–74.

Krehbiel, K. 1985. Obstruction and representativeness in legislatures. *American Journal of Political Science* 29:643–59.

Lebow, R. 2000. Contingency, catalysts and international system

change. *Political Science Quarterly* 115:591-616.
Lieberson, S. 1991. Small N's and big conclusions: an examination of the reasoning in comparative studies based on a small number of cases. *Social Forces* 70:307–20.
Lieberson, S. 1997. The big broad issues in society and social history: application of a probabilistic perspective. In V. McKim and S. Turner (eds.) *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences.* Notre Dame: Unversity of Notre Dame Press.
Lieberson, S. 2004. Comments on the use and utility of QCA. *Qualitative Methods: Newsletter of the American Political Science Association Organized Section on Qualitative Methods* 2:13–14.
Linz, J., and A. Stepan. 1996. *Problems of democratic transition and consolidation: Southern Europe, South America, and post-communist Europe.* Baltimore: Johns Hopkins University Press.
Lipset, S. 1959. Some social requisites of democracy: economic development and political legitimacy. *American Political Science Review* 53:69–105.
Little, D. 1991. *Varieties of social explanation: an introduction to the philosophy of social science.* Boulder: Westview Press.
Lupia, A., and M. McCubbins. 1998. *The democratic dilemma: can citizens learn what they need to know?* Cambridge: Cambridge University Press.
Mahoney, J. 2004. Comparative-historical methodology. *Annual Review of Sociology* 30:81–101.
Mahoney, J. 2004. Reflections on fuzzy-set/QCA. *Qualitative Methods: Newsletter of the American Political Science Association Organized Section on Qualitative Methods* 2:17–21.
Malinowski, B. 1944. *A scientific theory of culture and other essays.* Chapel Hill: University of North Carolina Press.
Moore, B. 1966. *The social origins of dictatorship and democracy: lord and peasant in the making of the modern world.* Boston: Beacon Press.
Morton, R. 1999. *Methods and models: a guide to the empirical analysis of formal models in political science.* Cambridge: Cambridge University Press.
Niou, E., P. Ordeshook, and G. Rose. 1989. *The balance of power: stability in international systems.* Cambridge: Cambridge University Press.
Organski, A., and J. Kugler. 1980. *The war ledger.* Chicago: University of Chicago Press.
Ostrom, E. 1991. *Governing the commons: the evolution of institutions for collective action.* Cambridge: Cambridge University Press.
Owens, D. 1992. *Causes and coincidences.* Cambridge: Cambridge University Press.
Pahre, R. 1997. Endogenous domestic institutions in two level games and parliamentary oversight of the European Union. *Journal of Conflict Resolution* 41:147–74.
Radcliffe-Brown, A. 1952. *Structure and function in primitive society: essays and addresses.* New York: Free Press.
Ragin, C. 1987. *The comparative method: moving beyond qualitative and quantitative strategies.* Berkeley: University of California Press.
Ragin, C. 2000. *Fuzzy-set social science.* Chicago: University of Chicago Press.
Richardson, L. 1960. *Arms and insecurity.* Pittsburgh: Boxwood Press.
Riker, W., and I. Sened. 1991. A political theory of the origin of property rights: airport slots. *American Journal of Political Science* 35:951–69.
Roth, A. 1995. Bargaining experiments. In J. Hagel and A. Roth (eds.) *The handbook of experimental economics.* Princeton: Princeton University Press.
Skocpol, T. 1979. *States and social revolutions: a comparative analysis of France, Russia, and China.* Cambridge: Cambridge University Press.
Tarrow, S. 1989. *Democracy and disorder: protest and politics in Italy, 1965–1975.* Oxford: Oxford University Press.
Taylor, M. 1987. *The possibility of cooperation.* Cambridge: Cambridge University Press.
Waldner, D. 1999. *State building and late development.* Ithaca: Cornell University Press.
Weldon, S. 2002. *Protest, policy, and the problem of violence against women: a cross-national comparison.* Pittsburgh: University of Pittsburgh Press.
Wohlforth, W. 1995. Correspondence: realism and the end of the Cold War. *International Security* 20:186–87.

It Ain't Necessarily So—Or Is It?

David Waldner
University of Virginia
daw4h@virginia.edu

Scholars who have couched their propositions in the form of necessary conditions owe a great debt to Gary Goertz for his recent work evaluating the status of claims to necessity and developing methods for testing them. In his most recent essay, Goertz makes a powerful case that necessary conditions hypotheses can reasonably be understood as probabilistic, not deterministic. Goertz suggests two benefits from this reconceptualization: necessary conditions hypotheses are more likely to get a fair hearing from social scientists who would otherwise dismiss them as deterministic, and those hypotheses will not be held hostage to single counter-examples that would unfairly be taken as disconfirming when in fact they likely represent measurement error.

But I think that the pragmatic benefits might be accompanied by some less-welcome implications. And I worry that switching to probabilistic language threatens connotative confusion. Taken together, these concerns make me reluctant to embrace Goertz's otherwise attractive position.

Define a necessary condition hypothesis as $P(X/Y) = 1$. Since P can vary between 0 and 1, it appears arbitrary to treat the range 0 to .99 as a homogeneous set contrasted sharply with $P = 1$. Why should the step from .99 to 1 constitute a threshold qualitatively different from the step between .98 to .99? Thus, Goertz argues, we suffer no loss of propositional meaning when we think of necessary conditions as "extremely likely" but not certain. Fine, but once we make necessity probabilistic, what prevents us from adopting, as Charles Ragin has advocated, other linguistic qualifiers, such as "usually necessary" or "necessary more often than not? As James Mahoney recently commented,

Empirically speaking, it is clearly useful to know if some

> factor X (or some value Z on variable X) is necessary or sufficient...90% of the time (or perhaps even 50% of the time). It is unclear why one would dismiss the accumulation of knowledge about these kinds of probabilistic causes on any substantive or policy grounds. (Mahoney, 2004: 84).

But why stop there? If stating that P = .99 permits us to claim that the relationship is "extremely likely," while P = .51 prompts us to claim that X is "necessary more often than not", then there is no non-arbitrary reason to prohibit me from claiming that when P = .25, X is "hardly ever necessary" for Y and when P = .01, X is "virtually never necessary" for Y. All of these would count as necessary conditions hypotheses.

These probabilistic expressions of necessary conditions violate ordinary usage. As my undergraduate logic textbook puts it, "if the event occurs every one of the conditions necessary for its occurrence must be fulfilled" (Copi, 1978: 284). To cite the example given in the *Oxford Dictionary of Philosophy*, "If p is a necessary condition of q, then q cannot be true unless p is true...Thus steering well is a necessary condition of driving well" (Blackburn, 1994: 73). I think most of us would be astonished to learn that "steering well is more often than not a necessary condition of driving well" or that "having gas in the car is usually a necessary condition for the car to run."

Indeed, it is unclear to me what knowledge would be gained by transforming deterministic necessary conditions into probabilistic ones. Ragin worries that randomness and error-riddled data conspire against finding clear evidence of necessity or sufficiency. (Ragin, 2000: 107-9). But if randomness is truly an inherent property of the social and political world, then, by necessity, necessity ain't. (Put differently, the absence of randomness is a necessary condition for the presence of necessity). And what exactly would we lose if we denied necessity? The position of Ragin and Mahoney implies that claims of necessity bestow inordinate epistemic value even when the meaning of necessity is being stretched beyond the normal usage of logicians and lay people. It is true that science has often valued invariance—but not invariably so. And besides, if the quest for invariance motivates striving to demonstrate necessity, what value can be gained by denying invariance via adding linguistic modifiers such as "usually?" Whatever value might be conferred by claims of necessity is adverbially eliminated. To see why this must be so, try to identify the difference between these two claims:

(1) X is almost always necessary for Y.
(2) X is almost always observed when Y is observed.

The claim that probabilistic necessity constitutes a clear gain of knowledge is based on the tacit assertion that something important is involved when we replace "almost always observed" with "almost always necessary." It might be argued that while (2) is cast in the language of association, (1) is cast in language that implies causation. But then the substitution is unwarranted and causal claims are being illicitly smuggled in. The problem with probabilistic necessity is that claims of necessity are based entirely on frequency distributions, not any knowledge of underlying causal processes and structures. In other words, when the proportion of cases displaying the proper values reaches some arbitrary threshold, the theory (the verbal summation of the distribution of data points) wins the label "necessary condition" along with the appropriate linguistic modifier.

I think it would be helpful at this point to distinguish two ways to think about necessary conditions:

(3) We never observe Y without also observing X; call this the frequency distribution approach.
(4) We provide reasons—causal mechanisms—why Y cannot occur without the prior occurrence of X; call this the generative approach.

The first formulation resonates with empiricist understandings of science: one examines the data and moves inductively to hypotheses. The latter formulation comports with a realist understanding of science and places greater emphasis on deductive theorizing and testing. In the example I use with my students: enrolling in my course is a necessary condition for receiving an A in the course. This is not a trivial condition: most enrolled students do not receive an A, and many students in my courses are auditors who receive no grade. An administrative rule functions as a causal process that helps generate the observed frequency distribution P(enrolled status/receiving an A) = 1. Knowledge of the rule itself permits us to treat the relationship as one of necessity; in this class of cases, we do not even need a frequency distribution to know that registration is necessary for the A. Indeed, from a causal standpoint, we might even argue that X is not a necessary condition for Y when P(X/Y) = 1. Consider this hypothetical example. Academic performance that meets my exacting standards is not a necessary condition to receive an A: bribery and threatened coercion are genuine substitutes, and so no amount of linguistic gymnastics can render meaningful the claim that "excellent performance" is a necessary condition to get an A. This remains true even though P(excellent academic performance/receiving an A) = 1: no student has yet attempted bribery or coercion. Yet by definition, even if ambient social processes permit only one causal pathway to be observed, when genuine causal substitutes exist, no single causal pathway is necessary. Frequency distributions in the absence of causal analysis cannot, therefore, subtend claims of necessity.

In short, we can think of three ways to establish the value of necessary conditions hypotheses:

(5) The property "necessity" is important, but only when not modified linguistically.
(6) The property "necessity" is important, even when modified linguistically.
(7) The property "probabilistic necessity" is an honorific that implies the virtues of invariance and causality without delivering them.

It seems to me that defenders of probabilistic necessity all wish to reject (5) and affirm (6), but they give no grounds to distinguish (6) from (7). Probabilistic necessity is, to conclude with a by-now overworked metaphor, the equivalent of grade inflation.

Goertz's essay focuses on causal explanations and mechanisms, signaling his sympathy with the generative approach. And Goertz asks only that we express necessary conditions hypotheses in probabilistic language: he says nothing about how to interpret them ontologically. Yet the language of probabilistic necessity tends to blur the boundaries between the frequency distribution and generative approaches and would, I think, be tacitly seen as endorsing the former.

Is this too strict a standard for testing? Goertz writes that one would not reject a necessary condition hypothesis just because a small minority of data points do not fit the model. Many would agree with Goertz that measurement error itself rules out the possibility of observing deterministic relations. I would urge caution before importing views of measurement error from quantitative to qualitative work. All too briefly, our understanding of measurement error is rooted in efforts to measure the true values of objective properties; in the 18th century, scientists who accepted Newton's deterministic framework were perplexed: when taking multiple measurements of a celestial object's location (an objective property) using better and better instruments and techniques, they still came up with different values. The great breakthrough occurred in 1808 when Carl Friedrich Gauss demonstrated that the distribution of random errors followed the normal curve. Measurement theory thus deals with measurement errors that are observer-independent, inevitable, and yet statistically corrigible. Many measurement errors in qualitative research are categorically different: they deal with the explication and application of concepts, and so they are observer dependent. (Searle, 1995) A consequence of observer-dependent errors, I would argue, is that while they are avoidable in principle, they are often intractable in practice. We need, in short, a long conversation about the nature of measurement error in qualitative research before qualitativists rely on measurement theory to relax their deterministic understanding of necessary conditions.

Finally, what should the defender of a necessary condition hypothesis do when confronted with noncompliant cases? Let me speak from my own experience, as a scholar who proposed necessary condition hypotheses based only on four cases and with limited knowledge of population parameters. I made deterministic claims because the reasoning underpinning them suggested determinism, not because I knew the larger frequency distribution. I made those claims in full knowledge that evidence from a greater number of cases would cast doubt on them. By making stark claims, I thought it would be even easier to identify discomforting, disconfirming evidence. When colleagues suggest flaws in my reasoning or counter-cases, I am forced to think long and hard about how to refine or reformulate my arguments. I fear that having probabilistic necessity at my disposal would be entirely too disarming; it would relieve me of the obligation to revisit my arguments. Goertz intends to ensure that necessary conditions hypotheses get a fair hearing; I worry that his approach might unintentionally breed complacency.

References

Blackburn, Simon. 1994. *The Oxford Dictionary of Philosophy*. Oxford University Press.

Copi, Irving M. 1978. *Introduction to Logic*. 5th Ed. MacMillan Publishing Co.

Mahoney, James. 2004. Comparative-Historical Methodology, *Annual Review of Sociology* 30: 81-101.

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

Searle, John R. 1995. *The Construction of Social Reality*. The Free Press.

Probability Tests Require Distributions[1]

Jasjeet S. Sekhon
Harvard University
jasjeet_sekhon@harvard.edu

I applaud Goertz's attempt at interpreting hypotheses about necessary conditions as probabilistic. As I discuss elsewhere in detail (Sekhon 2004), given the weakness of social science theories and the poor quality of our measurements, only probabilistic tests should be seriously considered. However, the move from deterministic theory to probabilistic testing is not a simple one. Although many difficult issues arise, I will here only discuss one: the issue of what distributions are required to conduct probabilistic tests and where they come from. Deterministic theories, by their very nature, do not say anything about the probability distributions in play. But the precise distributions used in any empirical test have profound implications. An interesting recent case of the confusion which can arise if one is not careful about this issue when comparing competing formal models is discussed and clarified by Wand (2005). I am also dubious that deterministic theories in general can be fixed *ex post* by adding stochastic terms instead of incorporating them from the beginning. For example, one has to be careful when simply adding stochastic terms that the theory remains self-consistent.

The problems associated with not explicitly incorporating stochastic terms into functional forms one wishes to test are apparent in the two examples Goertz offers:

$$Y = B_0 + B_1 X_1 \tag{1}$$

$$X \text{ is necessary for } Y \tag{2}$$

Goertz provides Figures 1 and 2 to show how each functional form fits the data. Given the information provided by him, it is not possible to conduct a statistical hypothesis test regarding the fit of either model without making additional assumptions. For example, Goertz offers the fact that the R^2 of the line in Figure 1 is 0.85 to support the contention that Equation (1) fits the data well. This is in itself irrelevant to the issue of whether the true relationship between X and Y is linear and stochastic. The fact that the R^2 is less than one does make clear, however, that it is false that the relationship between X and Y is linear and deterministic. But the issue of a stochastic linear relationship remains unresolved. For example, the data may be generated by the following process:

$$Y = B_0 + B_1X_1 + e \tag{3}$$

where e is distributed normally with mean zero and some variance σ^2. Unlike Equation (1), Equation (3) does provide enough information to be statistically tested in a variety of ways. But it is important to note that if we draw data from this model, we can by varying the value of σ^2 make the R^2 take any value between 0 and 1. Regardless of the value of σ^2, the functional form which maps X to Y would be linear, but the spread around the regression line defined by B_0 and B_1 would vary with the value of σ^2 (Achen 1982: 59). Moreover, without the assumption of normality (asymptotic or finite sample) there would be many questions about which loss function one should use to draw the regression line—i.e., the loss function one should use to choose the unknown parameters. For example, should one minimize the sum of squared residuals—$\sum_i (Y_i - \hat{Y}_i)^2$—which results in the usual least squares estimator? Or should one minimize the median of squares, which results in a robust estimator (Mebane and Sekhon 2004; Western 1995) that is useful if the distribution of E may be contaminated? There are many other alternatives, such as a regression assuming a t distribution. We cannot choose among the many alternatives in a principled fashion unless we make some distributional assumptions hopefully based on some knowledge about how the data were generated or collected.

Admittedly, too many quantitative analysts fail to justify the distributional assumptions made by the estimators they use. It is common to make vague gestures toward some central limit theorem and then quickly move on to the normality assumption. Although this behavior is predictable it is ironic given the great care which early users of regression took when defending their distributional assumptions (e.g., Yule 1899).

In any case, quantitative analysts usually either explicity or implicitly rely on limit theorems when justifying distributional choices. But it is unclear what one is to do with Equation (2). Given that the hypothesis is about necessary causation, why are three deviant observations not sufficient to reject the hypothesis? Is it because of measurement error or is it because the process is inherently stochastic? And what precise distributions are in play? Different distributions lead to different solutions and estimators. In other words, something specific needs to be said about the form of measurement error being proposed or the stochastic processes being considered. Of course, the same could be said about Equation (1) if it is not augmented in some fashion to incorporate some distribution.

It is interesting and potentially important that Goertz and others are trying to interpret necessary condition hypotheses in a probabilistic fashion. However, much work remains to be done in this fruitful area of research.

Notes

[1] I thank Jonathan Wand for comments. All errors are my responsibility.

References

Achen, Christopher. 1982. *Interpreting and Using Regression*. Newbury Park, CA: Sage Publications.

Mebane, Walter R. Jr. and Jasjeet S. Sekhon. 2004. "Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data." *American Journal of Political Science* 48 (2): 391–410.

Sekhon, Jasjeet S. 2004. "Quality Meets Quantity: Case Studies, Conditional Probability and Counterfactuals." *Perspectives on Politics* 2 (2): 281–293.

Wand, Jonathan. 2005. "Comparing Models of Strategic Choice: The role of uncertainty and signaling." *Political Analysis* 13 (3): TBA.

Western, Bruce. 1995. "Concepts and Suggestions for Robust Regression Analysis." *American Journal of Political Science* 39: 786–817.

Yule, Undy G. 1899. "An Investigation into the Causes of Changes in Pauperism in England, Chiefly During the Last Two Intercensal Decades (Part I.)." *Journal of the Royal Statistical Society* 62 (2): 249–295.

Announcements

2006 Award for Conceptual Innovation in Democratic Studies

The Committee on Concepts and Methods (C&M) of the International Political Science Association and the Centro de Investigación y Docencia Económicas (CIDE) in Mexico City invite submissions to the 2006 Award for Conceptual Innovation in Democratic Studies. The award is given every three years at the World Congress of the International Political Science Association (IPSA). The first was awarded at the 2003 IPSA World Congress in Durban, South Africa, to Gerardo L. Munck and Jay Verkuilen for their 2002 article "Conceptualizing and Measuring Democracy: Evaluating Alternative Indices" (*Comparative Political Studies* 35: 5–34). The second award, jointly sponsored by C&M and CIDE, will be given at the 2006 IPSA World Congress in Fukuoka, Japan.

The 2006 award will be given to a scholarly work published any time before December 31, 2005. Any category of formal publication may be submitted, whether it is a book, book chapter, or journal article. The idea of "conceptual innovation" is to be understood broadly; it is intended to cover concept analysis and concept formation as well as operationalization and measurement. The notion of "democratic studies" is to be comprehended widely, too; it includes research on authoritarian rule, democratization, and democratic quality. We seek to reward conceptual innovations that bear strong implications for empirical research.

Submissions are open to authors, journal editors, and book publishers. We encourage self-nominations. When submitting the work of others, please make sure you have obtained the express consent of the author. Board members of the Committee on Concepts and Methods are banned from participation.

Submissions must include: *four copies* of the work you submit, mailing address, phone, fax, and e-mail of the author, and mailing address, phone, fax, and e-mail of the person who submits. All submissions must reach us *before January 31, 2006*.

Please submit nominations by mail to:

Andreas Schedler, C&M Chair
CIDE Centro de Investigacion y Docencia Economicas
Division de Estudios Politicos
Carretera Mexico-Toluca 3655
Col. Lomas de Santa Fe, CP 01210 Mexico City, Mexico

The members of the 2006 jury are Robert E. Goodin (Australian National University, Canberra), Gerardo L. Munck (University of Southern California, Los Angeles) Chairperson; and Cindy Skach (Harvard University).

Fifth Annual Training Institute on Qualitative Research Methods, Arizona State University, January 3-14, 2006

The Consortium on Qualitative Research Methods (CQRM) is pleased to announce its fifth annual training institute on qualitative methods for graduate students and junior faculty.

The institute seeks to enable students to create and critique methodologically sophisticated qualitative research designs, including case studies, tests of necessity or sufficiency, and narrative or interpretive work. It will explore the techniques, uses, strengths, and limitations of these methods, while emphasizing their relationships with alternative approaches. Topics include research design, concept formation, methods of structured and focused comparisons of cases, typological theory, case selection, process tracing, comparative-historical analysis, congruence testing, path dependency, interpretivism, counterfactual analysis, interview and field research (including archival) techniques, necessary and sufficient conditions, fuzzy set methods, and philosophy of science issues relevant to qualitative research. Attendees will receive constructive feedback on their own qualitative research designs, and the course will also include master class discussions led by the authors of well-known works which employ qualitative methods. Examples will be drawn from exemplary research in international relations, comparative politics, and American politics. The syllabus from the fourth annual institute indicates the range of the issues to be covered. Please note, however, that this syllabus will be revised for the fifth institute, and should be viewed with this in mind.

CQRM member institutions will use their own meritocratic criteria to select students or junior faculty to attend the institute, and must notify CQRM of their choices by October 20, 2005. Students, fellows and junior faculty not so selected, or who attend non-member organizations, may apply directly to CQRM (see application form below). Applications must be received by October 20, 2005, and should consist of a curriculum vitae; a list of any courses taken in qualitative or other methodology; a short (300-word) personal statement briefly summarizing the applicant's main current research project and reasons for applying to the institute; and the name and contact information for a referee who is familiar with the applicant's training and research. Applicants will be notified of the outcome by November 11, 2005.

CQRM will cover the costs of tuition, lodging, and meals for successful applicants. Attendees will be responsible for their own transportation costs to and from Arizona State University. Participants for the institute will arrive on Monday, January 2, and depart late Saturday, January 14, or any time on Sunday, January 15. The seminar will meet daily, beginning on Tuesday, January 3. The final meeting is scheduled for Saturday, January 14.

Application form for the 2006 Institute on Qualitative Research Methods, offered by the Consortium on Qualitative Research Methods

Name____

Institutional Affiliation______________________________________

Address ____

____

____

Telephone: ____________________ Email: ______________________________

Discipline and Sub-field(s)____________________________________

WHAT DO I HAVE TO SEND? (1) This form, completed, or an email with the information requested above, (2) a curriculum vitae; (3) a list of any courses taken in qualitative or other methodology; (4) a brief (300-words) personal statement summarizing your main ongoing research project and indicating why you wish to attend the institute; and (5) the names and contact information for a referee who is familiar with your training and research. As noted below, all this information can be sent in hard copy or emailed.

WHO SHOULD COMPLETE THIS FORM? CQRM member organizations for the 2005-2006 academic year will have spaces reserved at the institute. Members will choose their nominees using their own selection procedures. Students and/or faculty who are nominated for those slots should not use this form. Interested applicants who are (a) from these institutions but are not so nominated or (b) who are from non-member institutions should use this form.

WHERE SHOULD I SEND THIS FORM? Please send application materials by October 20, 2005, to Colin Elman, Executive Director CQRM, c/o Political Science Department, Box 873902, Arizona State University, Tempe, AZ 85287-3902. Alternatively, you can email your application to consortium@asu.edu, and you can direct any questions about the institute to this same email address

European Consortium for Political Research

Call for Papers: The 3rd European Consortium for Political Research (ECPR) General Conference will be held in Budapest September 8-10, 2005. A full section, "Methodological Innovations and Dilemmas in Contemporary Political Research," (Section 20) is dedicated to methodological debates.

Panel headings (tentative):

20-1: Causality and Big, Slow-Moving, and Invisible Processes
20-2: Methodology in Political Science: Standards? What Standards?
20-3: Solving small-N problems by focusing on sub-national units
20-4: Empirical Implications of Theoretical Models (EITM) of Democratic Institutions
20-5: Looking at Methodological Questions Normatively
20-6: Innovations in (MV)QCA and Fuzzy-Set Applications
20-7: Beyond Regression? Predictive vs. Postdictive Models
20-8: Mixed Methods Designs: Advanced Issues
20-9: Outliers: Concepts, Treatments, and Uses in Different Methodological Approaches
20-10: Enlarging Our Toolbox: Modeling Strategies in Political Science

See more details, including panel abstracts, at: http://www.essex.ac.uk/ecpr/events_generalconference/budapest/section_list.aspx.

World Congress of the International Political Science Association

The Committee on Concepts and Methods is seeking paper proposals for the 20th World Congress of the International Political Science Association to be held in Fukuoka, Japan, on July 9-14, 2006. The committee is an open and plural platform of discussion on basic conceptual and methodological issues in political science. It strives to promote methodological discussion that takes seriously both concept analysis and qualitative methods. Proposals on any relevant topic are welcome. If you are interested in being a discussant or panel chair, let us know. Please email proposals and inquiries to Fred Schaffer (schaffer@mit.edu) by June 10, 2005. Information about the conference can be found at: http://www.fukuoka2006.com/en.

Qualitative
Methods
Boston University
Department of Political Science
232 Bay State Road
Boston, MA 02215
USA

Nonprofit Org.
U.S. Postage
PAID
TUSCON, AZ
Permit No. 271

Qualitative Methods is edited by John Gerring (tel: 617-353-2756, fax: 617-353-5508, e-mail: jgerring@bu.edu). The assistant editor is Joshua C. Yesnowitz (e-mail: jcyesnow@bu.edu). Published with financial assistance from the Consortium for Qualitative Research Methods (CQRM) and Boston University. Opinions do not represent the official position of CQRM or Boston University. After a one-year lag, past issues will be available to the general public on-line, free of charge, at http://www.asu.edu/clas/polisci/cqrm/QualitativeMethodsAPSA.html. Annual section dues are $8.00. You may join the section on-line (www.apsanet.org) or by phone (202-483-2512). Changes of address take place automatically when members change their address with APSA. Please do not send change-of-address information to the newsletter.