Fall 2009, Vol. 7, No. 2

Qualitative & Multi-Method Research

Newsletter of the
American Political Science Association
Organized Section for Qualitative and
Multi-Method Research

Contents

Letter from the Section President: Pluralism as a Hard Choice

Colin Elman
Syracuse University
celman@maxwell.syr.edu

I want to thank outgoing section President John Gerring for all of his efforts and achievements. John worked tirelessly on behalf of the section. Its overall good health—as reflected for example in the section's size and the range of its activities—is a tribute to his leadership. I also want to take this opportunity to thank Gary Goertz for his outstanding editorship of this newsletter, and the section's 2009 Division Chair, Rudy Sil, for constructing a first rate program.

As I begin my term as President, I would like to take a few lines to reflect on the section's epistemic diversity. Although this range is wholly welcome, it also lends itself to what might be described as a "prolate spheroid" problem. It is natural for people to see a particularly shaped ball as the one used for their own game—Americans naturally assume it is a football, as an Englishman I would suppose it was a rugby ball, and so on.

The common shape we see in qualitative research is close engagement with subjects and cases, and data collection initiated at extremely short range. This approach produces thick, rich, and heterogeneous data. These data are then used to advance a range of analytical goals (depending on the episteme in which the research is located) including interpretations, descriptive generalizations, and causal inferences.

I hope that it is uncontroversial to observe that the section includes qualitative researchers whose work is embedded in and justified by different epistemes, including within-case and small-n, interpretive, and statistical/regression approaches. Accordingly, while we may see people doing what we think of as "qualitative research," the meaning of what they are doing, and the rules they will be following, are different as we move from episteme to episteme.

There are at least three different ways to handle this diversity:

> Monism establishes a single episteme as the 'right' way. This claim might be based on a variety of arguments, including an assertion of foundational logic, a declared monopoly on the ability to answer privileged types of ques tions, or the predicate ability to deconstruct knowledge claims.

> Eclecticism proposes that epistemic differences need not matter. It leapfrogs over them to a problem-driven pragmatism—use what works, in whatever combination it works.
>
> Pluralism embraces diversity as a virtue, and seeks to find the limits of association and commensurability between several equally valid epistemes.

The section needs proponents of all three strategies taking part in the discussion. In particular, pluralists are needed to balance the monists' one-sized-fits-all, and the eclectics' rejection of foundations.

Without the check of pluralism, pursuing monism will be an unpleasant and ultimately self-defeating process. Whichever of the approaches emerges victorious from the struggle, the section will be impoverished by the absence of those it managed to exclude. As for eclecticism, it may be liberating for individuals to declare "badges, we don't need no stinking badges." But a strategy of decoupling methodology from epistemology simply delays the inevitable reckoning when lurking incommensurability surfaces.

While I believe that pluralism is a necessary part of the conversation, I acknowledge that it is also a hard choice. It requires learning about other research traditions, and being tolerant of differences. It mandates the sharing of scarce resources. It also means becoming more careful about overly broad claims, whether made in the positive ("this approach is the future of political science") or the negative ("scholars studying a few or single cases are historians not political scientists"). Innocuous in the context of a single episteme, these types of statements are staggeringly insensitive to other research traditions, and antithetical to a section that strives to be inclusive.

As the incoming President, I want to make the argument that the section's breadth is an opportunity to be embraced. It is not just a problem to be overcome by a retreat to homogeneity, or to be ignored by proceeding as if our differences do not matter.

Symposium: Cautionary Perspectives on Multi-Method Research

Is Multi-Method Research Really "Better"?

Amel Ahmed
University of Massachusetts, Amherst
aahmed@polisci.umass.edu

Rudra Sil
University of Pennsylvania
rudysil@sas.upenn.edu

Recent scholarship in political science attests to the rapid proliferation of approaches engaged in multi-method research (MMR), research that employs two or more methods selected from an array of qualitative, quantitative, and formal methods typically used in the social sciences.[1] The general notion that different types of methods can be employed to advance or test a particular theory is not in itself new. Multi-method approaches have long been a feature of social science research, taken up usually out of necessity (e.g. Jick 1979). Where data conducive to one method was not available, scholars would turn to another in order to fill the gap.

What is different about the more recent movement towards MMR is the extent to which the use of multiple methods is undertaken self-consciously by a *single* scholar in a *single* work in relation to a *single* research question, predicated on the assumption that the use of different methods will yield better results in addressing that question. Indeed, for some, MMR has come to represent not a pragmatic response to the complexity of a given problem but an end in itself and even a new universal standard for good scholarship. The method wars in political science have given way to an apparent consensus on the primacy of MMR as a way to achieve better results and overcome the limitations of particular methods.

This trend is most apparent in comparative politics where some of the field's leading figures have called for MMR as a means of overcoming the limitations of single-method research (SMR). According to George and Bennett (2005), the differences between different methodological approaches give them "complementary comparative advantages." Many have gone beyond this, however, to argue that MMR in fact represents a better research strategy, producing greater inference and more reliable findings than SMR (Brady and Collier 2004; Lieberman 2005; Laitin 2002; George and Bennett 2005).

While the increased appreciation of diverse methodological approaches is a welcome trend, there is much to be alarmed about in the recent rush to assert the primacy of MMR. What began as a movement to employ "all means necessary" in the service of problem-driven research is quickly turning into a new dogma that researchers *must, or ideally should,* incorporate "all means available" to validate their work. Practitioners and advocates of multi-method approaches explicitly argue that, all things being equal, multi-method research is better than single-method research. The goal of this essay is to assess this claim. While "better" can be understood in many different ways, our interest is particularly in the epistemological viability of the claim. Do we in fact learn more by combining different methods? Does MMR actually increase the number or strength of the inferences we can derive in the course of carrying out an individual research project? Can multi-method research actually increase the validity of findings? We are not concerned with the limitations or particular shortcomings evident in works using MMR. We are more interested in working towards an understanding of whether in principle MMR should be privileged over SMR at all times. In other words, is good

MMR always better than good SMR?

Our argument is straightforward: The claim that MMR is inherently better than SMR is built on the faulty premise that one method can offer external validity for the findings of another. Different methods can at best corroborate each other's findings, but this does not yield a more compelling inference. We do not know more or know better as a result of triangulating different methods because different methods rest upon incommensurable epistemological foundations that even the most heroic attempts at translation cannot overcome. Though combining methods may and often does produce good scholarship, we find that MMR holds the same epistemological status as separate projects addressing the same question, and that SMR is no less likely to produce good scholarship. The "goodness" of scholarship ultimately depends on the care and originality with which research is designed and executed, not on the number of methods that are deployed. Thus, although MMR is certainly valuable for social science research and should be welcomed as a part of a broad repertoire of methods available to scholars, there is no epistemologically sound reason to elevate it above others. Below, we make this case by examining some common forms of MMR and then considering some of the hidden costs associated with prodding individual scholars to adopt a strategy of MMR in the context of a single project.

Statistical Analysis and Varieties of Case-Study Approaches

Despite the great expectations of social scientists, practical applications of MMR have revealed the limitations of trying to use different methods to introduce measures of external validity. Because methods are premised on different epistemological commitments, they tend to employ different types of variables or mechanisms and privilege different dimensions of social reality. They also tend to focus on concepts that can be operationalized within certain boundary conditions and thus may not easily translate to other modes of inquiry premised on different sets of theoretical priors. For these reasons combining methods often results in findings that are incommensurable, frustrating efforts to offer external validity.

Take for example one of the most popular forms of MMR in use today: the combination of statistical and case study analysis. Proponents of such an approach argue that it offers greater analytical leverage, as one method compensates for the limitations of the other. Statistical analysis is limited in its ability to identify a causal mechanism as it tends to focus on causal effects.[2] Case study analysis, on the other hand, is well suited to identifying a causal mechanism, but is has limited generalizability. The combination of the two, it is argued, gives us the best of both worlds. Statistical analysis can be used to identify a general distribution of causal effects whereas the case studies may be employed for the purpose of identifying causal mechanisms and revealing separate links in a causal chain (Lieberman 2005; George and Bennett 2005).

First, it should also be noted that efforts to combine statistical and case-study analysis are only limited to certain kinds of case studies—those designed on the basis of an empiricist epistemology. Empiricism, which emphasizes the temporal priority of positive empirical observations and thus privileges inductive logics, provides a foundation for the probabilistic worldview of statistical analysis and is also consistent with the use of case studies for testing hypotheses or developing hypotheses that can be subject to quantitative tests. But, case studies have a variety of purposes. Some are constructed so as to support a more deductive orientation to theory building, as in the case of "analytic narratives" (Levi 2004) used in conjunction with game theoretic models. Others are designed on the basis of more hermeneutic or phenomenological approaches that stress the interpretation of meanings held by actors within distinctive contexts (Yanow 2006).

But, even with a common empiricist orientation, the claim that combining the two methods offers greater analytical leverage is difficult to sustain on epistemological grounds. This is because in moving from one mode of inquiry to another the basic conceptual categories will necessarily shift. McKeown (2004: 140–146) makes the point that the kind of inference that is privileged in the quantitative worldview cannot be conflated with the broader process of scientific inference which, in the case of a case study, focuses attention on a fundamentally different task: explaining how a set of initial conditions enable particular mechanisms to have particular effects in one or more contexts. Thus, in moving from large-N statistical analysis to case study analysis, the case study will by its very nature introduce variables not present in the statistical analysis (for example, variables that are not quantifiable but whose effects can be observed within a given context). It is precisely the depth of inquiry that enables case study analysis to identify causal mechanisms and reveal their effects in relation to a given outcome. However, it is also this feature that makes the two modes of inquiry incommensurable. They are effectively examining two different sets of variables. Thus in this scenario, the case study cannot be said to either confirm or falsify the finding of the statistical analysis.[3] Though it may offer a plausible story, at an epistemological level, it offers no corrective for the built-in limitations of the statistical analysis. In such combinations, the statistical analysis and case study analysis effectively represent independent (thought perhaps complementary) intellectual exercises. The findings of one cannot be said to validate the other.

Following the reverse sequence does not make this problem any less intractable. For example, if one begins with case study analysis to establish a causal mechanism and then tests the causal relationship with large-N statistical analysis, the latter cannot serve as a source of external validity for the former. As George and Bennett (2005: 138) note, causal mechanisms operate at the ontological level and can be neither conflated with, nor subsumed under, hypothesized causal effects. Statistical analysis can validate the relationship between hypothesized causal effects and generalize it across cases, but it cannot show that the causal mechanism found in the original case study analysis operates in the same manner and produces the same effect across different spatial and temporal contexts. Without case study analysis of each case, one cannot verify that the same causal mechanism is at work in all, let alone approximate the general size of its effect upon a given dependent variable. Both can provide independent analysis of related research questions, but one does not validate the other.

Formal Modeling and Empirical Analysis

Another popular mode of MMR features the combination of formal modeling with some form of empirical analysis. Proponents of such approaches also make the claim that the combination of the two methods produces greater insight, as one can be used to compensate for the limitations of the other. The formal model is thought to provide analytical rigor while the empirical analysis grounds the investigation in some social context. With this sort of combination the different methods tend to have somewhat more defined roles: the formal model is used for deductive theory-building and the empirical analysis is used for the purposes of illustration or theory-testing. However, the two sets of methodological operations involve quite different foundations. Even where the empirical analysis and formal model are both conceived of as broadly positivist enterprises, the former follows from empiricism and the latter from logicism (Shapiro and Wendt 2005). Using either case studies or statistical analysis in conjunction with the construction of a formal model would require completely ignoring the foundational principles on which the latter is built. The empirical analysis is also constructed specifically to test deductive models rather than to generate alternative causal stories since the causal structure and explanatory logic of the model depends more on axiomatic principles and internal consistency rather than on inferences from observed regularities.

In the case of "analytic narratives" (Bates et al. 1998), for example, the case-specific narratives certainly provide context, but the causal story does not emerge from this context, and the interpretations of the contexts are not evaluated against the strength of other interpretations of the same contexts. Moreover, since "theory means formal theory" (Bates et al. 1998: 3), there is the question of how the deductivist logic of formal theorizing can be meaningfully combined with the interpretive logic informing the construction of a context-bound narrative. While it is indeed admirable that scholars are able to demonstrate their use of both extensive-form game theory and case-specific research (sometimes involving fieldwork or archival research), the core causal logic of the explanation is given by the assumptions and logics built into a particular game-theoretic model rather than a balanced process of moving between theory and data (Sil 2000: 375). To be sure, empirical analysis can reveal flaws in an existing model and potentially inform a new model, but this can be done in differently designed SMR projects motivated by different objectives as part of our collective efforts to further knowledge. There is no inherent value to insisting that a single study incorporate both a formal model, premised on a logical positivist worldview and partial to deductive theorizing, and a case-specific narrative developed through interpretive methods designed to generate a deeper understanding of a given context rather than a more general explanation. Thus, while both game theory and case study serve important roles, those roles are distinct, each dictated by the nature of the methods combined, and each producing distinct, fundamentally incommensurable research products.

In the case of the marriage between statistical and formal approaches (e.g. Goldthorpe 1997), too, the use of one method to validate the findings of another is highly problematic. Although statistical analysis and formal modeling both proceed from a broadly positivist foundation, as noted above, the inductively oriented empiricism undergirding the former is fundamentally at odds with the deductively oriented logicism of the latter. The empirical analysis may corroborate the findings of the formal model by capturing the expected distribution of outcomes, but it cannot capture the effects of cognitive mechanisms that are predicated on the assumption of instrumental rationality.

Statistical analysis can determine a general pattern corroborating the findings of a formal model, and case study analysis may offer an illustration of a particular dynamic, but neither mode of inductive inquiry can offer external validation of a model deduced from a priori axioms. Again the problem of epistemological incommensurability presents an obstacle in the quest for external validation. Though considering empirical evidence maybe a useful heuristic for formal modeling, the juxtaposition of the two does not strengthen the inference we can draw from the former or the confidence we can have in a given model.

The Hidden Costs of MMR

The argument up to this point is *not* that there is no value in examining a substantive issue through the application of different methods. It is that, in the common forms of MMR considered above, what one can learn by juxtaposing two or more methods within a single research product is not fundamentally different from what we would learn from separate studies using different kinds of SMR to address the same question. This is not to say that the findings of a single research product featuring MMR are not useful. But, epistemologically speaking, there is no intrinsic gain from insisting on always triangulating different types of methods within a single approach rather than encouraging scholars interested in different methods to use those methods to explore the same substantive question as part of a larger collective effort to generate insights into the question. The combination of methods in a single study does not resolve the problem of epistemological incommensurability, and thus cannot eliminate the tradeoffs built into each of the methods employed at various stages of a multi-method project. Thus it cannot be said that we know more or know better when multiple methods are deployed.

This critique may ring hollow to some. Even if MMR is simply a juxtaposition of different intellectual projects within the same work, it might be argued, is this not in itself an advance over SMR? More significantly, is it not more efficient to have a single scholar generate findings using different methods when investigating the same sorts of substantive problem? Our conservative answer to both questions is: perhaps in some rare instances, but generally not.

Here, the main argument concerns the ways in which research is organized in the discipline writ large and the ways in which individual scholars conceive of their roles and contributions within the discipline. One must consider the amount of time and energy that is involved in MMR in terms of the methodological training, the fluidity in applying each of the methods, and the total time spent conducting research. Here, three possibilities exist. First, a scholar may end up taking the same amount of time to produce a single product using MMR as it would have taken to apply different methods in different research products. Second, time pressures related to grant deadlines, promotions, or other professional considerations may force researchers to spread a more limited amount of effort over different pieces of the research product. This may ultimately hurt the quality of the scholarship, producing thin case studies, shoddy datasets, and unsophisticated models hastily put together to round out a multi-method project. Third, it may actually end up taking *more* time as a single scholar shifts gears from one phase of a project featuring one method to a different phase using another method, especially if the scholar requires more "retraining" or "retooling" to effectively apply different methods. In all three scenarios, either some component of the research product will suffer, or a heavy burden will be placed on individual scholars.

Besides the costs to the individual researcher, there are considerable costs to the discipline as well. As researchers become more and more diversified in their methodological skills, they will likely become less and less specialized, spending less time on the approach they are most skilled at or most passionate about. This will have the inevitable effect of diluting the pool of expertise in the field and decreasing our collective efficiency as each of us feels compelled to maintain proficiency in the application of quite different types of methodological tools. Where there is a pool of labor available, it is not at all clear what is gained by making all of the members of that pool make the same kind of investment in gaining the same array of skills. If anything, the gains to efficiency from more specialized training and from iterated applications of the same method are lost. While specialization can certainly be taken too far, there is no reason to think that the entire discipline will gain by having all its members trained to do multi-method research.

Finally, there is a danger that the move towards MMR will result in decreased scrutiny of the core assumptions and epistemic commitments underlying the methods we choose to employ. Because MMR holds out the hope that by the very act of combining methods one can somehow overcome their individual limitations, it also holds the danger that scholars will feel justifiably absolved from questioning the foundations of the multiple methods they employ. This is perhaps the greatest danger of MMR: that we as a community of scholars will lose our critical eye towards methods and, in the process, lose our awareness of the fundamental challenges that have accompanied social scientific inquiry from its very inception.

Conclusion

The notion that MMR is always better than SMR assumes that if MMR is done well, one method can be used to validate the other, and a single project can generate more robust results or more compelling findings. The reality of social science research, however, is that all methods have limitations. This is no less true when they are deployed simultaneously in order to investigate a certain research problem. The problem of epistemological incommensurability that has long plagued exchanges across different research traditions is not being resolved by MMR; it is simply being transferred from the level of the discipline or subfield to the level of the individual scholar and research project. In this context, combining methods may produce complementary results, which may be valuable from a pragmatic point of view in generating support for a particular proposition. However, from an epistemological point of view, such combinations have the same structure and value as separate studies addressing the same research question. Putting the burden of producing these separate studies on a single scholar may not produce any gains to efficiency and may, in fact, come with some significant costs both for the scholar and for the discipline as a whole. In light of these concerns, MMR should be thought of not as the new gold-standard in research but as part of a diverse repertoire of methodological approaches that may be useful to some scholars depending on the nature of the research question—and on the preferred skill-sets and intellectual passions of the individual scholar.

Notes

[1] This essay is part of a longer article that is currently under preparation. It also builds on Amel Ahmed and Rudra Sil, "The Logic(s) of Inquiry: Reconsidering Multi-Method Approaches." Working Paper No. 16 of the Committee on Concepts and Methods, International Political Science Association (November 2008).

[2] King, Keohane, and Verba (1994, 85–87, 225–227) view causal effects as logically prior to and more reliable than unobservable mechanisms. For KKV, the value of mechanisms is limited to their ability to generate new observations that may influence the level of confidence in causal inferences. David Waldner (2007: 154) interprets this position as flawed in that it fails to recognize the epistemological function of mechanisms, reducing them to mere "servants of inferences."

[3] Moreover, as explained by Rohlfing (2008) in the cases of "nested analysis," the introduction of case studies introduces bias potentially exaggerating findings.

References

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

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

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

Goldthorpe, John. 1997. "Current Issues in Comparative Methodology." *Comparative Social Research* 16, 1–26.

Jick, Todd. 1979. "Mixing Qualitative and Quantitative Methods: Triangulation in Action." *Administrative Science Quarterly* 24:4, 602–611.

King, Gary, Robert Keohane, and Sidney Verba. 1994. *Designing Social Inquiry*. Princeton: Princeton University Press.

Laitin, David. 2002. "Comparative Politics: The State of the Subdiscipline." In *Political Science: The State of the Discipline*. Ira Katz-

nelson and Helen Milner, eds. (New York: Norton), 630–659.
Lieberman, Evan. 2005. "Nested Analysis as a Mixed-Method Strategy in Comparative Politics." *American Political Science Review* 99:3, 435–452.
Levi, Margaret. 2004. "An Analytic Narrative Approach to Puzzles and Problems." In *Problems and Methods in the Study of Politics*. Ian Shapiro, Rogers Smith and Tarek Masoud, eds. (New York: Cambridge University Press), 201–226.
McKeown, Timothy. 2004. "Case Studies and the Limits of the Quantitative Worldview." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Henry Brady and David Collier, eds. 139–167.
Rohlfing, Ingo. 2008. "What You See and What You Get: Pitfalls and Principles of Nested Analysis in Comparative Research." *Comparative Political Studies* 41:11, 1492–1514.
Shapiro, Ian and Alexander Wendt. 2005. "The Difference That Realism Makes: Social Science and the Politics of Consent." In *The Flight From Reality in the Human Sciences*. (Princeton: Princeton University Press), 19–50.
Sil, Rudra. 2000. "The Foundations of Eclecticism: The Epistemological Status of Agency, Culture, and Structure in Social Theory." *Journal of Theoretical Politics* 12:3, 353–387.
Waldner, David. 2007. "Transforming Inferences Into Explanations: Lessons from the Study of Mass Extinctions." In *Theory and Evidence in Comparative Politics and International Relations*. Mark Lichbach and Richard Ned Lebow, eds. (New York: Palgrave), 145–176.
Yanow, Dvora. 2006. "Thinking Interpretively: Philosophical Presuppositions and the Human Sciences." In *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*. Dvora Yanow and Peregrine Schwartz-Shea, eds. (Armonk, NY: M. E. Sharpe), 5–26.

The Challenge of Conceptual Stretching in Multi-Method Research

Ariel I. Ahram
University of Oklahoma
arielahram@ou.edu

Multi-method research (MMR) has gained enthusiastic support among political scientists in recent years. Much of the impetus for MMR has been based on the seemingly intuitive logic of convergent triangulation: two tests are better than one, since a hypothesis that had survived a series of tests with different methods would be regarded as more valid than a hypothesis tested only a single method. In their seminal *Designing Social Inquiry*, King, Keohane, and Verba (1994) argue that combining qualitative and quantitative methods is useful because it increases the amount of data used to test a specific theory or hypothesis. While critical of specific prescriptions in KKV, Brady and Collier (2004) and Gerring (2007, 2009) reiterate the mantra of epistemological monism, shared standards, and logical consistency between qualitative and quantitative methods.

This paper, though, warns that what Sartori (1970) calls the "stretching" or "straining" of concepts between qualitative and quantitative domains has potentially damning implications on MMR. Simply because qualitative and quantitative findings point in the same direction—statistical significance and coefficient signs match the outcome of a case study—does not make them any more likely to be true, since the concepts applied in one methodological component are not equivalent to those applied in the other. It is impossible for qualitative and quantitative methods *to say the same thing* because they are talking *about different things*. An alternative schema, though, is possible based on seeing the component of an MMR design as representing two distinct cultures of inquiry that are complementary, rather than corroborating.

Conceptual Stretching and Causal Analysis

The topic of conceptual stretching was first broached as a warning against the proliferation of quantitative, statistical methods and the incessant drive to substitute quantifiable variables in place of qualitative categories. Conceptual stretching occurs, according to Sartori, when "*denotation is extended by obfuscating the connotation*" (Sartori 1970: 1041). The term itself is used in the context of the ladder of abstraction, in which concepts are mapped along two inversely-related dimensions: the *intension* (connotation), the systematic and explicit definition of the characteristics of the concepts, and the *extension* (denotation), the range of cases which can be categorized as meeting the conceptual definition. Increasing a concept's extension by incorporating more cases under its rubric leads to stripping away some of its necessary intension, the specificity of characteristic involved in the concept. While ascending or descending the ladder is critical for theory building, Sartori deplores attempts to expand a concept's extension without acknowledging concomitant diminishment in intention. Collier points out some exceptions to the law of inverse relation, but maintains the fundamental argument that in travelling to cover new cases, conceptual definitions can suffer unacknowledged distortion (Collier and Mahon 1993; Adcock and Collier 2001).[1] Such distortion makes a concept's terms, definitions, and referents inconsistent, violating a crucial criterion in evaluating social scientific work (Gerring 2009: 112).

To understand the nature of conceptual stretching in MMR better, a closer comparison of qualitative and quantitative approaches to conceptualization is necessary. Coppedge (1999), an early and eloquent advocate of MMR, describes qualitative concepts as "thick," having complex definitions developed iteratively through examination of a small realm of cases. In contrast, quantitative concepts are "thin," with relatively simple conceptual definitions. Conceptual thickness/thinness is inversely related to narrowness/breadth of extension. Because of their definitional intricacy and high intension, qualitative concepts are designed to apply to only a small number of cases. The proliferation of vocabulary about various democratic and authoritarian regime types and subtypes is exemplary of the type of highly descriptive conceptual categorization used in qualitative analysis. Quantitative scholars, by comparison, rely on datasets like Freedom House or POLITY, which reveal only that two countries are equally democratic (or undemocratic) and have no substantive meaning to the distance between intervals (Munck and Verkuilen

2002). On the other hand, the simpler definitions and low intension involved in quantitative concepts are amenable to incorporating much wider universe of cases for interrogation via statistical analysis. Thus, concepts like development as measured in per capita GDP can be applied widely and easily analyzed in terms of scalar relationships of cases (Ragin 2000).

Fundamentally, the difference between qualitative and quantitative approaches to measurement is ontological. Since measurement is about relationships of objects to one another, a change in measurement will lead to a different set of relationships (Franzosi 2004: 281). Qualitative measurement necessarily involves the use of *categorizing* cases using a specific nominal definition. For such a schema to be meaningful, membership must be absolute and each case must be equivalent in the same set of crucial characteristics. The goal is to eliminate measurement error by devising a comprehensive conceptual framework in a clearly delineated case domain. By contrast, quantitative measurement involves scoring cases so that every case is related to the other on an interval scale (Mahoney 2000; Goertz 2008).

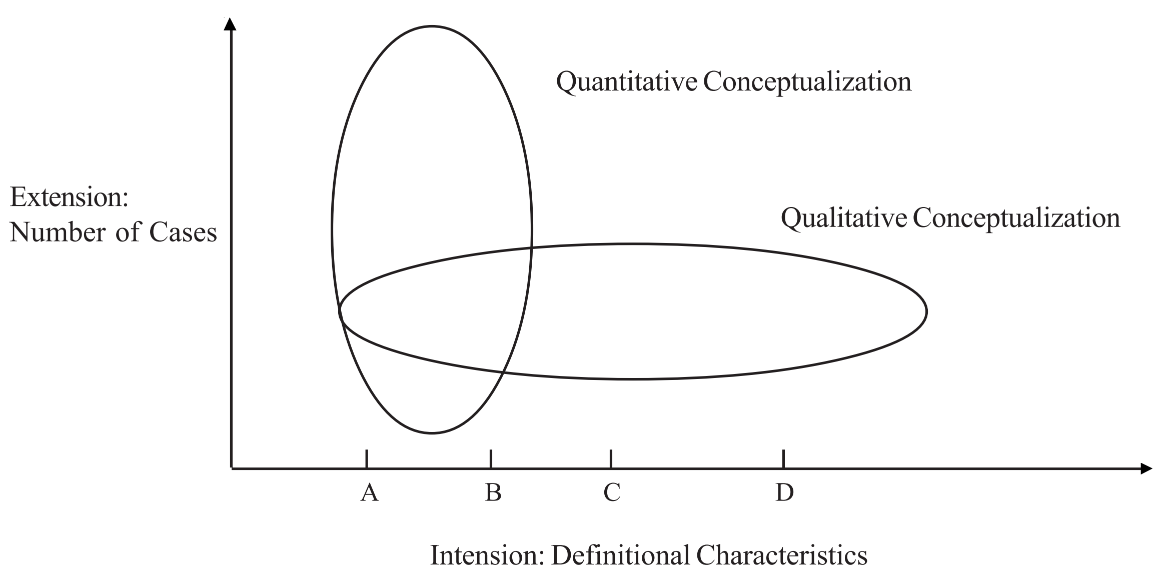
To some, the danger of measurement error and problems of causal inference which can arise when quantitative indexes do not relate closely to a given concept is more apparent than real. After all, qualitative concepts can be represented in a quantitative equation using categorical (i.e., "dummy") variables. Still, there are several problems with this solution: the more complex a conceptual definition becomes, the more difficult it becomes to summarize using quantitative measurement across a large sample and the more tenous the crucial assumption of unit homogeneity becomes (Gerring 2007: 52). In a statistical sense, the use of dummy variables represents a significant sacrifice in analytical power and cannot be arrayed in a meaningful interval format (Taagepara 2008: 58, 299–231). Moreover, 'translation' between qualitative and quantitative data is not transitive: while qualitative data might travel into a quantitative domain, the opposite translation is difficult, if not impossible (Creswell and Clark 2006: 128).

Moving between qualitative to quantitative concepts, then, is bound to cause some distortion or mismatch in the intension and extension in qualitative and quantitative concepts. While concepts may share a label or term, they have different characteristics and refer to different sets of cases with varying attributes. This is depicted in Figure 1 below where the horizontal axis lists four characteristics necessary for falling under the concepts rubric (A, B, C, and D) and the vertical axis the range of empirical cases that fall under it.

In the figure, conceptual stretching is most obviously manifest at the level of extension: numerous cases that would be counted as "in" in a quantitative setting but "out" in a qualitative setting. But a more troubling slippage occurs at the intensive dimension, where characteristics that are crucial in the qualitative definition of a concept are absent in the quantitative one. As Goertz notes, causal hypotheses are embedded within a concept's connotation. In the qualitative setting, specific causal hypothesis are dependent on cases having attributes A, B, C, and D. These attributes explicate the mechanisms by which causes produce their effects (Goertz 2006). Cases that do not share these attributes, such as those in quantitative domain that share only A and B, cannot be considered candidates for at least some of the mechanisms hypothesized in the qualitative domain.

This insight is crucial whether a variable appears as explanandum or explanan. For example, consider Sen's critique of quantitative, income-based headcount conceptualizations of poverty and argument for thickening the concept by considering social exclusion and relative inequality (Sen 1987). This revision changes who is counted as poor and therefore forces an explanation of poverty that accounts for a different universe of incidents. If the category of the "impoverished" includes victims of social exclusion who might still earn above

Figure 1: Overlay of Thick and Thin Conceptualization



a given standard of living, then a range of solely economic policy prescriptions are inadequate to effectuate change. Alternatively, the addition of these social dimensions to the concept of poverty also has implications for studies of the poverty's *effects* on outcomes such as democratization, terrorism, or civil war, where the definition of the dependent variable is often belabored but the independent variable left merely as dichotomous.[2] Conceptualization is always implicitly a process of defining a scope condition, delineating a set of cases with common attributes which are hypothesized to have the same etiological properties. If one or more of these attributes is present in the qualitative conceptualization but absent in the quantitative conceptualization, this circumscribes the range of properties to which causation can be attributed (Lucas 2003; Goertz 2006).

Conceptual Stretching in Practice

Conceptual stretching is a potential challenge in all social scientific work, but it takes particular forms in MMR. One of the most obvious manifestations is alterations between the use of nominal and ordinal modes to represent same variable. For instance, it is logically problematic to claim that five cases are in the set of democracies and then to claim that one of these is "more democratic" than the others. In some cases this is the result of imprecise or sloppy language. But such inconsistency obfuscates what precisely the key variables are meant to account for and the relationship of the cases to one another (Mahoney 2000: 408).

A more subtle form of stretching can be seen in the difficulties of establishing qualitative and quantitative case domains, since case selection is intimately tied to conceptualization (Collier and Mahoney 1996). Evan Lieberman's *Race and Regionalism in the Politics of Taxation in Brazil and South Africa* (2003) exemplifies a methodologically astute and sophisticated work that nonetheless exhibits unacknowledged conceptual stretching between qualitative and quantitative renderings of the critical variable. Lieberman's central contention is that

> [t]he specification of group rights in the form of official state documents and policies provides a stronger set of incentives for political entrepreneurs to make claims based on such identities… Federalism, for example, tends to give important political salience to regional identities, and official racial exclusion tends to give much more salience to racial identities… (Lieberman 2003: 14).

Privileging certain types of identity over another, which Lieberman calls the concept of *national political community* (NPC), constrains the ability of political entrepreneurs to demand the creation of progressive taxation systems.

Lieberman's qualitative, historical investigation of Brazil and South Africa comes to a counterintuitive but persuasive conclusion: South Africa's 1909 Constitution specified white supremacy while denying recognition to regional differences. By establishing whites as a legal category, this cornerstone document encouraged the white economic elite to cooperate with the state in establishing a social safety-net system that raised the living standards of their poorer co-ethnics. After apartheid's downfall this redistributive system was opened to all races, leading to greater economic equity overall. In contrast, Brazil's 1891 federalist constitution privileged claims based on regional equity, but was explicitly inclusive on racial grounds. Faced with the prospect of racial equality, Brazil's white economic elite opposed the state's effort to develop a redistributive system that would benefit a mainly black underclass.

The stretching of NPC becomes apparent as Lieberman attempts to test whether similar racial and regional definitions of NPC the same effect in other cases. Examining constitutions and other legal documents from over one hundred cases, he converts the data into a series of dummy variables. The statistical results show a correlation consistent with the small-N study, but the regression model does not capture important dimensions of the qualitative findings. Specifically, the historical narrative focuses on the interplay of racial and regional identity, but in the quantitative analysis these variables are not interacted. More importantly, the qualitatively-derived elaboration of NPC proves too narrow to incorporate the majority of empirically relevant cases. The initial, thick-version of NPC is limited to two dimensions and four different modes, as shown in Table 1. It is the concept's thickness that allows Lieberman to make specific hypothesis about the identity basis for collective action.

Table 1: Distribution of Cases in Lieberman's Qualitative Analysis (2003: 79)

	Federal	Unitary
Race Exclusionary		South Africa
Race Inclusionary	Brazil	

Table 2: Distribution of Cases in Lieberman's Quantitative Analysis (2003: 242)

		Federal	Unitary
Fragmented	Race Exclusionary	3	4
	Race Inclusionary	2	10
Not Fragmented		7	43

But in the larger dataset of sixty-nine countries, forty-three cases in the population (62%) have neither relevant racial or regional cleavage that are crucial to the initial definition of NPC and have to be incorporated under residual category of "non-fragmented" states, as shown in Table 2. In the absence of significant racial or regional cleavages, Lieberman's qualitatively-derived theory about collective action is mute and the mechanisms he posits for explaining collective action cannot account for the majority of the cases in the expanded conceptual domain.

Complementarity Triangulation as an Alternative Logic of MMR

Recent proponents of MMR focus assume the easy commensurability of concepts and focus mainly on whether or not quantitative or qualitative results confirm one another. If the findings of one method do not "fit" the initial model, then its hypotheses are rejected and specification amended (Lieberman 2005; Fearon and Laitin 2008). A case in point is the Political Instability Taskforce, which "having at least two independent approaches [one qualitative, one quantitative] to assessing instability, if they point in the same direction, greatly increases the confidence of predictions" (Goldstone 2008: 7). While Rohlfing (2008) cautions that this approach is susceptible to error because qualitative and quantitative technique might have compounding rather than correcting biases, this essentially adopts the monist assumption of comparability between methods. The problem of conceptual stretching—the mismatch between concepts and variables using in different settings—is more fundamental because it questions the very comparability and compatibility of qualitative and quantitative methods. If what is categorized in a qualitative domain as instability is not equivalent to what is scored as such in a quantitative dataset, agreement can be dismissed as mere felicitious coincidence.

An alternative logic to MMR, however, seeks to leverage the complementary aspect of qualitative and quantitative methods. Such an approach assumes that qualitative and quantitative research methods construct different objects, do not measure the same objects, or focus on different attributes and properties of the object, which in sum make mutual validation impossible. Rather,

> [t]he linkage of quantitative and qualitative methods does not produce one unique picture of reality. Instead, the different research results would have to be combined in a sort of jigsaw puzzle to produce an adequate image of reality. (Erzberger and Prein 1997: 144)

The fitting-together of this puzzle depends crucially on the differing theoretical underpinnings of the use of each method, particularly their ability to offer explanations at different levels of analysis. Each method offers a unique capacity for explanation. Contrary to the monists, Mahoney and Goertz (2006) argue that qualitative and quantitative methods represent distinct cultures of inquiry, making inherent some friction between the two. Specifically, the use of nominal measurements in qualitative methods is tied to an approach which focuses on the "causes of effects" in individual cases. In contrast, the use of interval measurements in quantitative methods is tied to an approach focused on estimating the average effect of independent causes. These traditions each offer their own standards by which to evaluate causal claims, but these standards are in a sense complementary:

> an explanation of an outcome in one or a small number of cases leads one to wonder if the same factors are at work when a broader understanding of scope is adopted, stimulating a larger-N analysis in which the goal is to explain particular cases and more to estimate average effects. Likewise, when the statistical results about the effects of causes are reported, it seems natural to ask if these results make sense in terms of the history of individual cases, one wishes to try to locate the effects in specific cases (231).

An example of this type of cross-cultural dialogue comes from multi-volume collaborative study of civil wars by Collier and Sambanis (2005). This work uses case studies by individual country experts to explore a statistical model that analyzes the impact of opportunities to seize property ("greed") or capitalize on thwarted political ambitions ("grievances") on civil war onset. In reviewing the contribution of the qualitative component, Sambanis (2004) notes that several of case studies identified country-year intervals as periods of civil wars ("1" in the dichotomous dependent variable) which were coded as non-war ("0") in the dataset. This is not simply an example of measurement error in need of correction. Rather, it goes fundamentally to the connotation inherent to the concept of civil war itself, which has ramifications for the logit/probit results obtained in the statistical analysis. Similar conceptual revisions were made of independent variables. Statistical evidence shows that states with lower levels of education have a greater likelihood of suffering from civil war, since citizens with little education are more likely to be able to "do well" by war by engaging in criminal behavior during civil war. But some countries with relatively well-school citizens, like Cyprus, Yugoslavia, Georgia, Russia, and Lebanon, experienced civil war, while Saudi Arabia, with a low education rate, has not. Qualitative studies suggest that it is not just the *extent* of schooling but the *type* of education which can influence the propensity for war. A curriculum rife with ethnic chauvinism, for instance, can also reinforce the motive for violence. Combining a conceptualization of education as consisting of both qualitative categorical and quantitative interval components opens up the possibility of new types of explanations for the apparent linkage of antecedent and outcomes.

The power of such complementary triangulation comes not in testing and rejecting specific hypothesis, but in refining concepts and thereby the range mechanisms that account for the outcome of interest. Areas of congruence and incongruence between qualitative and quantitative concepts must be investigated inductively. Dunning's (2007) advice that a researcher code a handful of cases from a pre-packaged dataset manually to ensure familiarity with variable features and underlying conceptual definitions is doubly important in MMR. The most relevant cases for the purpose of conceptual clarification are the most liminal at the independent or dependent variables. Determining the disposition of cross-over cases—whether (or not) they can be considered democracies or to have experienced civil war—is crucial because these cases highlight precisely the criterion of inclusion and exclusion within a categorical set (Goertz 2008; Ragin 2000). The goal is not to amend qualitative and quantitative conceptual definitions until they are equivalent, but to identify their divergences.

Explicating empirical heterogeneity and conceptual incongruity is vital to capturing the multiple and manifest pathways connecting structural antecedents and outcomes. Like all uses of narrative techniques to trace causal mechanisms, MMR must establish its truth claims by testing alternative explanations within a single case (George and Bennett 2005). But identifying a single mechanistic process is rarely enough to account for the full range of correlation observed in the quantitative analysis. In the example from Lieberman above, the problem is not that the mechanisms specified by qualitative methods are incompatible with the statistical pattern. Rather, the problem is that the mechanisms are logically precluded from functioning in many of the cases within the large-N population. Put another way, a significant portion of the R^2 observed in the quantitative portion must have been due to mechanisms other than the ones identified in the study. More attention to the exact parameters of stretching between qualitative and quantitative concepts allows the researcher to look for these alternative mechanisms.

Since inferences about mechanisms are derived from the properties of the concept used in the qualitative analysis, the power of MMR is constrained by the representativeness of the qualitative sample. Qualitative concepts and methods are the fulcrum, highlighting heterogeneity among case units and locating multiple mechanistic pathways connecting initial conditions to outcomes. The success of complementary MMR hinges on exploring the overlap between large and small-N conceptualization. Thick and thin concepts should not be rendered equivalent by assumption or theoretical deduction. Rather, inductive techniques must be used to demonstrate the contours and boundaries of their shared analytic space.

Notes

[1] For more on the relationship between intention and extension, see Goertz (2006: Ch. 3).

[2] See also, Brady (2003).

References

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

Brady, David. 2003. "Rethinking the Sociological Measurement of Poverty." *Social Forces* 81:3, 715–752.

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

Collier, David and James E. Mahon. 1993. "Conceptual 'Stretching' Revisited: Adapting Categories in Comparative Analysis." *American Political Science Review* 87:4, 845–855.

Collier, David and James Mahoney. 1996. "Insights and Pitfalls: Selection Bias in Qualitative Research." *World Politics* 49:1, 56–91.

Collier, Paul and Nicholas Sambanis, eds. 2005. *Understanding Civil War: Evidence and Analysis*. Washington DC: World Bank.

Coppedge, Michael. 1999. "Thickening Thin Concepts and Theories: Combining Large N and Small in Comparative Politics." *Comparative Politics* 31:4, 465–476.

Creswell, John W. and Vicki L. Plano Clark. 2006. *Designing and Conducting Mixed Method Research*. Thousand Oaks, CA: Sage.

Dunning, Thad. 2007. "The Role of Iteration in Multi-Method Research." *Qualitative Methods* (Spring), 22–24.

Erzberger, Christian and Gerald Prein. "Triangulation: Validity and Empirically-Based Hypothesis Construction." *Quality & Quantity* 31:2, 141–154.

Fearon, James and David Laitin. 2008. "Integrating Qualitative and Quantitative Methods." In *Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry Brady, and David Collier, eds. (New York: Oxford University Press), 756–778.

Franzosi, Roberto. 2004. *From Words to Numbers: Narrative, Data, and Social Science*. New York: Cambridge University Press.

Gerring, John. 2007. *Case Study Research: Principles and Practices*. New York: Cambridge University Press.

Gerring, John. 2009. *Social Science Methodology: A Criterial Framework*. 2nd Edition, New York: Cambridge University Press.

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

Goertz, Gary. 2006. *Social Science Concepts: A User's Guide*. Princeton: Princeton University Press.

Goertz, Gary. 2008. "Concepts, Theories, and Numbers: A Checklist for Constructing, Evaluating, and Using Concepts or Quantitative Measures." In *Oxford Handbook of Political Methodology.* Janet Box-Steffensmeier, Henry Brady, and David Collier, eds. (New York: Oxford University Press), 97–118.

Goldstone, Jack A. 2008. "Using Quantitative and Qualitative Models to Forecast Instability." Washington DC: U.S. Institute of Peace.

King, Gary, Robert Keohane, and Sideny Verba. 1994. *Designing Social Inquiry*. Princeton: Princeton University Press.

Lieberman, Evan S. 2003. *Race and Regionalism in the Politics of Taxation in Brazil and South Africa*. New York: Cambridge University Press.

Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99:3, 435–452.

Lucas, Jeffrey W. 2003. "Theory-Testing, Generalization, and the Problem of External Validity." *Sociological Theory* 21:3, 236–253.

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

Mahoney, James and Gary Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14:3, 1–23.

Munck, Gerardo and Jay Verkuilen. 2002. "Conceptualizing and Measuring Democracy: Evaluating Alternative Indices." *Comparative Political Studies* 35:1, 5–34.

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

Rohlfing, Ingo. 2008. "What You See and What You Get: Pitfalls and Principles of Nested Analysis in Comparative Politics." *Comparative Political Studies* 41:11, 1492–1514.

Sambanis, Nicholas. 2004. "Using Case Studies to Expand Economic Models of Civil War." *Perspectives on Politics* 2:2, 260–279.

Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64:4, 1033–1053.

Sen, Amartya. 1987. *The Standard of Living*. New York: Cambridge University Press.

Taagepera, Rein. 2008. *Making Social Sciences More Scientific: The Need for Predictive Models*. New York: Oxford University Press.

Ontology, Epistemology, and Multi-Methods

Abhishek Chatterjee
University of Virginia
ac7y@virginia.edu

Enthusiasm for multi-methods research can possibly be ascribed to the prima facie promise it holds for moving beyond, if not resolving, seemingly intractable debates on the relative merits of "qualitative" (historical, interpretive, etc.) versus "quantitative" (i.e. inferential statistical) research methods.[1] The justification of multi-methods rests on the claim that combining a few case studies with a larger *inferential*—and not descriptive—statistical study manages to capture the strengths of both insofar as the discovery of causal relations is concerned. This in turn lends greater confidence that the relationships being asserted are indeed causal. The specific argument is that since inferential statistics allows for generalization (while case studies normally do not), and case studies are better at tracing what are called "causal mechanisms," combining the two affords us the best of both worlds.

The trouble with this is that scholars seeking to justify multi-methods seem to assume that the question of what constitutes "cause" or "causal mechanism" is unproblematic, and the problem is limited to that of making causal claims. The problem, in this view, is solely epistemological. Epistemologies however do not exist in vacuum; they are both supported by and in turn support ontologies (or metaphysics), which can roughly be defined as presuppositions or innate conceptions about the nature of the world. An insufficient appreciation of this leads to mutually contradictory arguments in favor of multi-method research designs; arguments, which on reflection could not possibly support such designs. Arguments conceding the usual weaknesses of case studies—but nonetheless attempting to justify them—imply a metaphysics that makes it impossible to portray case studies as either necessary or sufficient in causal analysis, which in turn also precludes any justification of multi-method research. In other words, some fundamental concessions—implicitly based on a specific ontology—negate almost all subsequent justifications that could be made in favor of case studies, and by extension, multi-method designs.

The causal ontology often accepted in pointing out the deficiencies of case studies—implicitly or explicitly—is "reductionist" and "regularist," i.e. one which respectively defines causes in terms of non-causal relations and states of affair and affirms that such non-causal relations are regularities in nature. The origins of this metaphysical view can be traced to David Hume (1999 [1748]: 136)—hence often referred to as "Humean."[2] The particular conception of what it means to make 'causal generalizations' is a logical implication of this ontology of causality. Moreover the idea of 'generalization' cannot be separated from the definition of causality here; in other words to say that something is caused by something else is also to generalize in a certain way, namely, by referring to regularities. Though inferential statistics finds sufficient justification in (and in turn sufficiently justifies) this ontology of causality, explanations based on case studies are not consistent with it. Case studies and inferential statistics cannot logically mix if the definition of causality is reductionist and regularist. This also applies to arguments claiming that case studies illuminate causal mechanisms, since the only definition of "mechanism" that is consistent with this ontology is one that sees them as concatenation of variables that occur with some regularity, something that case studies are not equipped to handle. Multi-methods using case studies can therefore never be justified under this metaphysical view.

Yet (1) referring to regularities is not the only way to generalize, (2) causes do not necessarily have to contain generalizations, and (3) it may not be possible to reduce causes to something more basic. In each of these three cases, one can find sufficient justification for case studies (and also *independently* for inferential statistics), but the usual arguments for combining the two run into logical difficulties. This is because the usual justification for multi-method designs is in fact a confusion of distinct metaphysical views about the nature of causation that are not necessarily complementary. How does one know that the mechanism connecting a cause with an effect in a particular case study is the same mechanism connecting causes to effects in all the other cases? What part of the study does the causal work, the case studies or the statistical analysis? If it is the case study then the statistical analysis should not convince us, and if it is the statistical analysis then the case study should not convince us. This epistemological dilemma arises because the problem is not merely methodological; it involves our fundamental, and most often implicit, metaphysical assumptions about the nature of the world. Let us examine these issues in turn.

That small-N is not merely an epistemological problem becomes evident when we ask under what definition of "cause" should small-N be a problem for establishing causal relations. The answer has to do with statistical theory and the Humean conception of causation that sufficiently—though not necessarily—justifies it. To understand this, let us consider the epistemological and methodological implications of this conception. In other words, given a Humean view, how would one go about discovering causal relations? Now very briefly, Humean definitions come in both deterministic and stochastic versions. Causes precede their effects, and are either necessary, sufficient, or both necessary and sufficient conditions (in the deterministic versions), or increase the conditional probability of their effects (in probabilistic versions). In both cases, every singular causal statement must be an instance of one of more general causal laws. The singular phenomenon itself need not be repeated as long as the unique phenomenon can be shown to be the result of a combination of laws that recur in other singular phenomena. Epistemologically therefore, the singular phenomenon cannot play a role in the establishment of a causal relationship since it is itself dependent on preexisting regularities that have already been established. Both the deductive nomological (D-N) scheme of explanation,proposed most clearly by Hempel and Oppenheim (1948), and Hempel's (1942) inductive-statistical (I-S) scheme follows directly from such conceptions of causation.

The epistemological problem is that of discovering regularities when many laws are instantiated simultaneously. Under ideal conditions experimentation would be the first best method (this obviously is not unique to Humean views; experimentation as a method is consistent with almost all ontologies of causation, but interpretations of experiments would differ depending on the definition of causality). One way to overcome the problem of simultaneous instantiation would be to isolate individual causes and observe their effects repeatedly to establish lawlike regularities. When we move from the experimental sciences to the social or non-experimental sciences, the goal remains the same, i.e. the discovery of regularities, but this time they have to be detected from purely observational data. This is where statistical models come in. Such models try to approximate the experimental situations described above. These models assume that the data being generated are akin to the result of a series of independent experiments or observations generated from mutually independent processes where nature manipulates the independent or explanatory variable under different background conditions or controls (again, it is also possible to give other *interpretations* to inferential statistics). Inferential statistics is also consistent with the definition of causes as generalizations; that is, the "regularity" part of the definition, or alternatively the definition of causes as "types." The latter is obviously because insofar as it informs one of average effects, generalization (over a particular population) is built into the interpretation of inferential statistics.

The link between a reductionist and regularist metaphysics on one hand and inferential statistical methods on the other should be clearer now. The impossibility of fitting case studies into this framework should also be evident. Indeed, some prior discussions in political science have clearly recognized this. For example, Sartori defended comparative case studies as a third-best method behind experiments and statistical studies (1994:16). His argument was that though it is true when it comes to drawing causal inferences, comparative case studies are inferior to either experiments or statistical control, the phenomena that most interest certain political scientists do not occur enough times to lend themselves to statistical studies. The problem with this defense is that the acceptance of the logic of statistical inference entails that a few cases cannot or should not lead us to believe that a cause exists. This is the crux of Lieberson's (1991) argument against drawing causal conclusions from a few comparative cases (also see Sekhon, 2004). Using the example of automobile accidents, Lieberson shows how fragile our conclusions can be as to the causes of accidents if we rely on only a few cases, assuming that knowledge of causes entails knowledge of regularities. The most logical conclusion in this instance would be to state that given the paucity of cases one cannot say anything about the presence or absence of causes.

Lieberson's critique applies equally to solutions to the problem that urge us to somehow increase the number of cases by, among other things, performing "within case analyses" by looking at multiple implications or consequences of a particular theory or causal statement within the same case (Campbell 1975: 184–189). But if we assume that regularities are most basic and knowledge of causes entails knowledge of regularities, it is difficult to count multiple implications as an augmentation of the number of cases. For at a given level of analysis, each implication of any causal statement must be considered separately. It is for this reason that statistical models require each observation to be independent. And multiple implications of the same causal statement or theory can never be considered independent from each other. There is a rebuttal to Lieberson's argument, but as we shall shortly see, it makes sense only within decidedly non-Humean ontology of causation. Within the Humean ontology, Lieberson's position is very convincing indeed.

Again, early discussions seemed to have conceded this. Still case studies were defended variously as "a first stage of research, in which hypotheses are carefully formulated," (Lijphart 1971: 685), or as explications of particular cases for their own sake in light of theory, as in Verba's "disciplined configurative approach," (1967: 114–115) among others. In such an approach the researcher seeks to explain the event with the help of established regularities and general causal statements. It is important to note here that though disciplined configurative explanations rest on general laws, the explanation itself does very little to strengthen or weaken the validity of the said laws (Lijphart 1971: 692). Yet these concessions are sometimes accompanied with arguments that cannot easily be reconciled with the former. Thus, for example, Lijphart's subsequent assertion that such studies can be considered "crucial experiments" if values on the variables are extreme is difficult to reconcile with his statement quoted above. Why should extreme values on variables in one case cause us to reexamine our prior theory, especially since the latter could be based on a large number of cases? The same applies for 'deviant case' analyses. As Mckeown (1999) has also observed in a slightly different context, a single additional case can never, by this logic, lead us to weaken an original proposition that is, in Lijphart's own words, "solidly based on a large number of cases" (1971: 692). The problem is that some of Lijphart's epistemological points about the contributions of case studies make sense only when decoupled from his ontological orientation which seems to underlie the bulk of his other points.

Another popular defense of case studies—that such studies are better at handling determinism (Gerring 2004: 347, Munck 1998: 33)—is based on conflation of ontology with epistemology. It is perfectly consistent to have a deterministic and Humean view of causality—indeed, the original Humean view was in fact deterministic and some philosophers have argued that "Hume and indeterminism don't mix" (Dupre and Cartwright 1988)—and still claim at the epistemological level that statistical inference is the best way to establish this causality. As Laplace observed a long time ago, an (ontologically) deterministic relation can appear to be (epistemologically) stochastic because of ignorance of all relevant laws and initial conditions. It does not matter whether the view of causality is deterministic or stochastic as long as causation is reduced, and it is reduced to regularities either deterministic or stochastic. In both situations case studies can never be logically justified as the "first best" method. Similar interpretations can also be given to the use of inferential statistics in political science. As a result, criticisms such as Lieberson's against the use of Mill's methods would still be valid. What we discover from Mill's methods cannot even be considered "cause" in this sense.

The reference above was to deterministic sufficient conditions. But can deterministic necessary conditions justify case studies, as Dion (1998) has argued? Dion's argument protects case studies against the small-N criticism only under extremely restrictive conditions. The argument has more to do with the problems that classical inferential statistics faces in tackling necessary conditions than the inherent strengths of case studies. In fact it could be seen primarily as an advocacy of Bayesian statistics over classical statistics when it comes to necessary conditions.

Since Bayes' rule depends crucially on known probabilities to determine posterior probabilities, its applicability is limited to only certain kinds of systems. To be precise, it is crucial that the mechanism that generates prior probabilities is well-known, and alternative hypotheses have well-known probability outcomes or likelihoods. The prior probability is a source of great debate in both philosophy and statistics (See Sober 2002, for example). It is uncontroversial in cases of systems where there is a clear way of assigning prior probabilities. But it is slightly more controversial in cases where we can't. Then the question becomes what should the prior probabilities be based on? Should they be based on statistical regularities, "common sense," case studies, or subjective opinions? As soon as we ask these questions, we realize that we are back to the square one. Additionally, and more pertinent to the use of such statistics to defend small-N's is the fact that we would have to consider multiple hypotheses with determinate likelihoods for effective statistical control; at which point the difference in terms of sample size between classical inferential statistics and Bayesianism begins to disappear. Even this argument, as a result, cannot provide sufficient justification for case studies.

This brings us to the final and most popular set of justifications for both case studies and their incorporation in multi-methods research, namely, that case studies are uniquely suited to discover or enunciate what are called "causal mechanisms," which statistical studies are less able to do. However, "mechanism" is yet undefined. Further, of two possible understandings of the concept (of mechanism), one does not provide any justification for case studies, while the other—while sufficiently justifying case studies—cannot easily support their incorporation in multi-method designs.

If mechanisms are defined as, "in effect, variables that operate in sequence," (Sambanis 2004: 288), or any variation thereof, some of the same criticisms that we started with apply. The difficulty of defending case studies while holding this particular understanding of mechanisms stems from the fact that that it implies just another version of the Humean definition extended to intervening variables. It is theoretically possible to multiply the number of steps between cause and effect while remaining steadfastly Humean. Each link or mechanism in a longer chain can be represented by equations that can be construed as statements of regularity and as such the same epistemological concerns that were raised earlier about the confirmation of causal claims with case studies apply here too. Various statistical models such as path models would seem to be the natural recourse. If this is a fair representation of some definitions of causal mechanisms, then again the sufficiency of case studies cannot be defended.

More avenues open up once we abandon either reductionist or regularist (or both) understandings of the concept of "mechanism." But these latter conceptions, though equally supportive of inferential statistics independently, cannot easily accommodate the usual manner of performing multi-method research without running into logical contradictions.

"Singularist" definitions of causality hold that singular events and not regularities are more basic. The definition decouples generalizations from the definition of causality (Ducasse 1993; Salmon 1980, 1997). Epistemologically, therefore, one need not look for generalities, and the explanation of a single event or case can count as a causal explanation. Process tracing in case studies receives sufficient metaphysical justification here. But this ontology presents us with a problem. Such reductionist but singularist definitions of causality have difficulty distinguishing spurious causes from "real" causes at the definitional level. One way of overcoming this is to attach counterfactuals to singularist mechanisms. Counterfactuals, however, are very sensitive to contrast spaces. The truth condition of counterfactuals depends on the contrast space of any explanation and therefore causality also becomes context and contrast space dependent in this case. So, for instance, causes of revolutions as opposed to near-revolutions can be very different from causes of revolutions as opposed to non-revolutions, or revolution in country A as opposed to revolution in country B. Generalizations, if any, in this case are "bottom up" and change based on the relevant contrast spaces rather than "top-down" and ostensibly universal. Furthermore since singular events are more basic, there is no expectation that generalizations will necessarily emerge. But if we define contrast spaces with as much generality as possible, for instance in our example above, as all possible cases of near-revolution, and if we call answers to both kinds of questions (the limited and expanded contrasts, respectively) "cause," certain problems recur at the epistemological level in combining methods since there is no presumption that an answer to one question will have any bearing on an answer to the other. Thus though attaching counterfactuals to singular mechanisms suffices to justify case studies, they cannot justify multi-method research.

For instance Evan Lieberman's (2005) latest attempt to suggest a framework for multi-method research faces this particular problem. He writes that "a nested research design implies that scholars will pose questions in forms such as "What causes social revolutions?," while simultaneously asking questions such as 'What was the cause of social revolution in France?'" (2005: 436) For an answer to both questions to qualify as "causes" almost necessarily implies a singularist view of causation. Under a regularist view an answer to the second question cannot differ from an answer to the first, and the former has to be at least a subset of the latter. His advice is to start with a large-N analysis and then—in case of robust and satisfactory results—"test" the model with small-N analysis by choosing cases that fall within the average prediction of the large-N model (2005: 437). Why should we expect the small-N cases to be consistent with the large-N predictions? Even if they are, why should we have any confidence that the average prediction of the large-N analysis and case study research point to the same causal relationship? In the absence of robustness Lieberman advises model building and analysis of predictions that fall in the average, and also the outliers (2005: 439–440). The criteria for "robustness" and "satisfaction" must be statistical; it is therefore difficult to see why lack of robustness should motivate case studies. Indeed there are well-known remedies within inferential statistics for such problems as lack of statistical significance or any bias in a model and none of these involve looking at case studies. Note that all the questions raised here do not imply that Lieberman is wrong, but that the argument contains large gaps, owing to insufficient appreciation of the metaphysical implications of methods. Additional arguments have to be supplied to reconcile mixing of the two methods.

Case studies also receive sufficient justification if we abandon a reductionist view of causation or causal mechanisms. This would reverse the order of priority in the relationship between regularities and causes. Instead of regularities being signifiers or definers of causes, prior knowledge of causes would restrict and inform the kinds of inferences one is able to make from statistical relations. This is also an effective rejoinder to Lieberson's criticism of the comparative method. This is part of Nancy Cartwright's argument for considering causal "capacities" as primitive. She contends that it is the arrangement of capacities in certain ways that produce regularities; "nomological machines," or "socio-economic machines" as she calls them, are particular arrangements of capacities that "in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behavior that we represent in our scientific laws." (1999: 50) Capacities, further, cannot be identified by any particular manifestation. They can be compared to qualities such as kindness or tenacity that are carried by human beings. Such qualities are not identified with any one particular behavior; instead they are instantiated in multiple circumstances as different behaviors all of which have in common the fact that they are displays of kindness or tenacity (1999: 51). Socioeconomic machines are essentially fables that illuminate important aspects of how the world works, while capacities can be equated with morals of such fables. The relationship between the fable and the moral is not that of similarity but "that of the general to the more specific…[e]ach particular is a case of the general under which it falls" (1999: 39). This means inter alia that "satisfying the associated concrete description that applies on a particular occasion is what satisfying the abstract description consists of in that occasion" (Ibid). Thus any particular arrangement of capacities is also general, and in turn, every general capacity finds its manifestation only in particular arrangements. Once we understand capacities well enough (as is the case in certain natural sciences) we can further manipulate these capacities and arrange them in different ways to produce different laws. As Cartwright observes, "anything *can* cause anything else. In fact, it seems…not implausible to think that, with the right kind of nomological machine, almost anything can *necessitate* anything else" (1999: 72).

The epistemological consequences of this view urge us to treat both (most) large-N statistical studies and case studies as essentially alike in that both can be interpreted as attempting to 'guess' the arrangement of hypothesized capacities in the world. Sometimes when we know about enough capacities and other background conditions "[w]e accept laws on apparently slim experimental bases…[and] the data plus the description of the experimental set-up deductively imply the law to be established" (Cartwright 1999: 93). Case studies, both single and comparative, can therefore be considered similar to fables that substantiate morals. The fables however have to be very carefully constructed with great attention to capacities and their arrangements. They are necessarily concrete, but they are at the same time general. This is precisely why studies like John Gaventa's (1980) of one particular locality in one country are also general. Notice that domain restriction finds its best justification under this ontology. In fact if we follow this logic, restrictions of domains is imperative, since what we are describing are particular nomological machines, the very definitions of which carry the connotation of restriction. This is because as we observed earlier, it is the arrangement of particular capacities in certain orders and under certain conditions that could generate laws. But domain restriction does not mean restricted generalization. The fact that some physical laws are literally true only within the confines of the laboratory does not prevent them from also being general. This answers certain criticisms of case studies based on their domain restriction. Thus to say that domain restriction in case studies necessarily implies limited causal force is to implicitly accept an ontology that cannot justify case studies in the first place.

Even in this case, however, the usual manner of combining case studies with a large-N (inferential) statistical analysis cannot be logically supported because of the reasons pointed out earlier. On the other hand, one way of avoiding the usual contradictions in mixing methods would be to truly "triangulate" within the general framework of a case study. In other words instead of using the usual procedures of picking one case out of any sample, one could try to empirically describe the arrangement of capacities (of course, in the context of prior background knowledge of capacities) of any one case, and then examine the implications of such an arrangement using quantitative evidence. This would work because as Cartwright pointed out, it is the particular arrangement of capacities that produces regularities. But it must be a necessary preliminary to first explain why and how the arrangement of capacities came to be. This kind of suggestion is most relevant to the literature on institutions in political science and sociology, especially the ones based on single cases.

Notes

[1] Another possible but independent reason, particularly of interest to those interested the sociology of knowledge is that multimethod research, especially when used in doctoral theses signals to potential employers competency in both statistical and others kinds of research methods thus satisfying the largest possible coalition of potential employers. To reiterate, this is one possible hypothesis in need of further study, and will not be addressed further in this contribution.

[2] Though modern versions are significantly different from what Hume originally may have suggested.

References

Campbell, Donald T. 1975. "'Degrees of Freedom' and the Case Study." *Comparative Political Studies* 8:2, 178–193.
Cartwright, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
Dion, Douglas. 1998. "Evidence and Inference in the Comparative Case Study." *Comparative Politics* 30:2, 127–145.
Ducasse, C.J. 1993. "On the Nature and the Observability of the Causal Relation." In *Causation*. Ernest Sosa and Michael Tooley, eds. (New York: Oxford University Press), 125–136.
Dupre, John and Nancy Cartwright. 1988. "Probability and Causality: Why Hume and Indeterminism Don't Mix." *Nous* 22:4, 521–536.
Gaventa, John. 1980. *Power and Powerlessness: Quiescence and Rebellion in the Appalachian Valley.* Urbana: University of Illinois Press.
Gerring, John. 2004. "What is a Case Study and What is it Good For?" *American Political Science Review* 98:2, 341–354.
Hempel, Carl G. 1942. "The Function of General Laws in History." *Journal of Philosophy* 39, 35–48.
Hempel, Carl G. and Paul Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15, 135–175.
Hume, David. 1999 [1748]. *An Enquiry Concerning Human Understanding*. Tom L. Beauchamp, ed. (New York: Oxford University Press).
Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99:3, 435–452.
Lieberson, Stanley. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70:2, 307–320.
Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65:3, 682–693.
McKeown, Timothy J. 1999. "Case Studies and the Statistical Worldview." *International Organization* 53:1, 161–190.
Munck, Gerardo L. 1998. "Canons of Research Design in Qualitative Analysis." *Studies in Comparative International Development* 33:3, 18–45.
Salmon, Wesley. 1980. "Causality: Production and Propagation." PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association.
Salmon, Wesley. 1997. *Causality and Explanation*. Oxford: Oxford University Press.
Sambanis, Nicholas. 2004. "Using Case Studies to Expand Economic Models of Civil War." *Perspectives On Politics* 2:2, 259–279.
Sartori, Giovanni. 1994. "Compare Why and How: Comparing, Miscomparing and the Comparative Method." In *Comparing Nations: Concepts, Strategies, Substance*. Mattei Dogan and Ali Kazancigil, eds. (Oxford: Wiley-Blackwell), 14–34.
Sober, Elliott. 2002. "Bayesianism—Its Scope and Limits." In *Bayes' Theorem*. Richard Swinburne, ed. (Oxford: Oxford University Press), 21–38.
Sekhon, Jasjeet S. 2004. "Quality Meets Quantity: Case Studies, Conditional Probability, and Counterfactuals." *Perspectives on Politics* 2:2, 281–293.
Verba, Sidney. 1967. "Some Dilemmas in Comparative Research." *World Politics* 20:1, 111–127.

Speedbumps on the Road to Multi-Method Consensus in Comparative Politics

Michael Coppedge
University of Notre Dame
coppedge.1@nd.edu

Is there a multimethod consensus in comparative politics? My short answer is: not quite. For example, recently I was updating my department's reading list on research methods for the comprehensive exam in comparative politics, and I added a chapter by Lakatos (Lakatos 1970) to it and sent it to my colleagues for feedback. One replied, "I'm especially glad to see Lakatos added!" Another replied, "What is Lakatos doing in there?" (Actually, my colleagues are unusually collegial.) But there is very little consensus on any aspect of comparative politics, so it is unrealistic to expect anything resembling consensus in our subfield (España-Nájera, Márquez, and Vasquez 2003).

My longer answer is that there is rough agreement in principle that multimethod work would be a good thing. There is also agreement that, in practice, some aspects of multimethod work are hard to pull off. But I think that there are some other challenges in multimethod work that are not yet sufficiently appreciated—speedbumps on the road to the great multimethod harmonic convergence.

On the encouraging side, we agree that we can do case studies to verify causal mechanisms or explore anomalies identified by statistical analyses or formal theories; we can do statistical analyses to test whether arguments generated by case studies or formal theories are generally true; we can develop formal theories to explain tendencies turned up by case studies or statistical analyses; and so on, with many variations (Lieberman 2005).

On the discouraging side, we can agree that it is hard for any one researcher to develop cutting-edge expertise in all three methods, and that not being on the cutting edge can be an obstacle to publishing multimethod work. One can overcome this obstacle by collaborating with those who have greater expertise in different methods, but each person tends to feel that he or she is doing more work and getting only partial credit for it. And there is some truth to that (Bennett and Braumoeller 2009). These difficulties are well known and accepted. But I think there are other obstacles to multimethod work that have not received as much attention, and yet remain serious obstacles. The first is the mismatch between concepts used by different approaches. The second is disagreement about what to test and when to test. The third is the need for clarity about specifying appropriate boundary conditions.

To begin, there is a flaw in any multimethod work that relies on mismatched concepts—which probably includes most multimethod research. Different approaches have a strong tendency to use radically different concepts. Case studies employ thick concepts—complex, nuanced, multifaceted concepts, often even concepts with proper names; and quantitative analysis and formal modeling almost always employ thin concepts. There is a lot of information in any concept of "Silvio Berlusconi" that is not contained in any definition of "prime minister," much less "head of government," which are concepts more typically used in quantitative analysis; and these contain more information than "agent," "formateur," "chair," or whatever the closest corresponding concept would be in a formal model. Silvio Berlusconi is indeed a prime minister and a formateur of cabinets; but he is not *just* a prime minister and formateur.

This raises questions about whether it is safe to take an explanation of Italian outcomes and test it more generally with evidence about Gordon Brown, Stephen Harper, Manmohan Singh, or other prime ministers. Maybe it is, maybe it's not; it all depends on whether the characteristics of Berlusconi that are relevant for the test are ones that are shared with all the other prime ministers rather than any of his more colorful or scandalous characteristics. This difficult translation applies in the other direction as well. A hypothesis that is deduced to be true of formateurs under a stylized set of assumptions may not apply to actual prime ministers in general, or to Silvio Berlusconi in particular. Some of the stylized assumptions may not be true for them, and they may possess different motives, powers, or constraints that override the assumptions of the model. The mismatch between the concepts calls the relevance of the test into question.

These are major conceptual leaps that must be made carefully. We have to make sure that concepts that are used in one approach are the functional equivalents of concepts used in another approach if we are to get much benefit from combining the approaches. This means that as we climb the ladder of abstraction, we must leave behind the attributes that are irrelevant and take with us all the attributes that matter for the theory at hand. Unfortunately, knowing which attributes matter is hard. It requires round after round of theorizing and systematic testing.

My hope is that eventually we will develop a set of standardized concepts that can be used safely in any approach—concepts that have been repeatedly shown to be useful because they (a) identify consequential attributes of the political world and (b) have robust and plausible causal relationships with other useful concepts. I am not advocating the creation of a central authority that certifies some concepts and indicators as useful and bans others, however. The kind of standardization I have in mind would emerge naturally from the competitive efforts of many scholars and the judgment of their peers, who will slowly come to agree on the most useful ways to define and measure common concepts in comparative politics.

The second problem is that there is still considerable disagreement about what to test and when to test. From the deductive side, there are some who believe that the current emphasis on testing gets in the way of the use of model-building to provide insights, generate surprising hypotheses, and explicate causal processes. Clarke and Primo, for example, frustrated by the higher status accorded models that generate testable predictions, ask, "Why test predictions from a deductive, and thus truth-preserving, system? What can be learned from such a test? If a prediction is not confirmed, are assumptions already known to be false to blame?" (Clarke and Primo 2007: 741). Their answer is that models serve other valuable purposes—providing insights, organizing empirical generalizations, generating surprising hypotheses, defining causal mechanisms, and forecasting the future—and should not usually be subjected to testing (743).

From the inductive side, there are those who believe that it is not necessary to test conclusions that come from intimate engagement with detailed, rich historical, or case study evidence. There is a feeling in some quarters that if you know your cases really well, it is not necessary to see whether your conclusions hold up with systematic, fresh evidence. This stance is most evident in comparative historical research, in which shuttling back and forth between theory and evidence, which entails amending or supplementing the theory in the light of new evidence, is praised as a methodological virtue even though it blurs the line between theory development and testing (Mahoney 2003).

I object to both attitudes. I do respect the need to develop theories and models, drawing on both deductive and inductive approaches. This is a valuable, indispensable stage of the scientific process. But at some point, theories and models have to be tested systematically with fresh evidence, even if there is a presumption of truth arising from inductive engagement with the facts, and especially if they preserve the "truth" of assumptions that are known to be false. Otherwise, deductive approaches risk developing ever-more complicated models that have no connection to reality, and inductive approaches risk giving up on developing theoretical understanding altogether. I get annoyed when either formal modelers or comparative historians claim to have achieved "insights," to have made "progress," or to have "valuable" findings before anything has been subjected to a well-designed systematic test. Making good time on the highway is not progress if you're driving in the wrong direction. There have to be reality checks along the way. Testing is the equivalent of stopping to check the map or ask for directions from time to time.

Testing does not necessarily mean running regressions. That is certainly one kind of testing—*extensive* testing—which is appropriate when the goal is to generalize, whether to the globe or to the many observations to be found within a single country. But there is also *intensive* testing that tests which of several hypotheses is most consistent with many kinds of evidence from a single case. I hope that scholars will increasingly recognize that these are two equally valid, yet quite distinct, modes of testing. Neither do I mean to imply that everyone must test, all the time. Clearly there are benefits to having some division of labor. As a practical matter, it is important for some research programs to develop without testing for a while. But I do think that no project should go too far before *someone* does some testing (Green and Shapiro 1994).

I also think that we should make some effort to set off in directions that are likely to be right. That means, as much as possible, starting out with assumptions that are fairly realistic. Of course, no assumption is entirely true. All models simplify reality. That's what models do. But there is a crucial difference between assumptions that are reasonable simplifications and assumptions that are patently false. Reasonable simplifications are thought to be true more often than not; they may ignore minor complications, but they identify the most consequential features of reality. False assumptions may be correct in some situations—a stopped watch is right twice a day—but usually they are incorrect. They may focus on conditions that matter at the margins, but they ignore the most important determinants. I fail to understand what useful insights can emerge from assumptions that are known to be false. Clarke and Primo are right about that much: passing a test says nothing about the truth of a model if its assumptions are false.[1] However, this is not likely to happen unless the testing is flawed. Reasonable simplifications are more likely to lead to predictions that survive rigorous testing.

A good motto would be: Test Everything! Test assumptions, test predictions, test indicators for validity and reliability. As we test, however, we need to explicitly specify appropriate boundary conditions. The third obstacle to multimethod consensus is the fact that we are not in the habit of doing this. Yet we must: the political world is a heterogeneous place. There are not likely to be many one-size-fits-all propositions that hold true for all times and places. I think the kind of theory we are developing is a kind of hierarchical theory: there is one model that works in context A, a different model that works in context B; then a meta-model that explains why each model works in each context; and so on, with models nested inside meta-models nested inside meta-meta-models, which taken together constitute a theory. This is very similar to Bennett and George's notion of "typological theory" (George and Bennett 2005).

If this is correct, then I see two tendencies that inhibit the development of this kind of hierarchical theory. One is acting as though we did not live in a political world of complex conditionality. No one really believes this: that there are any non-trivial statements about politics that are true anywhere and anywhen. Nevertheless, when we test propositions by building databases that indiscriminately pool absolutely all the observations that we can find, we are acting as though there really is unit homogeneity: the same causes have the same effects in all possible contexts. To correct this tendency, we need to think very carefully about the non-arbitrary conditions under which our hypotheses should be expected to hold. We must also state these boundary conditions explicitly.

The opposite tendency is to take it for granted that everything is so complex and conditional that there is no point in trying to generalize. This leads to a narrowing of theory or even deemphasis of theory, and to myopia: the tendency to ignore large, slow-moving, structural causes and to give too much weight to small, dynamic, proximate causes. The remedy for myopia is exactly the same: thinking very carefully about the conditions under which our hypotheses should be expected to hold, and stating these boundary conditions explicitly. When we do this, the boundary conditions give us the cues we need to link the findings of case studies with the findings of large-sample analysis.

All of these problems are difficult because they involve breaking old habits, developing new skills, or exposing our cherished ideas to great risk of falsification. Nevertheless, I am optimistic about the future of multimethod research. It has real advantages and it is, I think, our future. All of these problems are in principle solvable. Being aware of them is the first step toward a better understanding of politics.

Note

[1] Clarke and Primo (2007) would object to the very notion of a model being true or not. They hold that models are merely objects; they are neither true nor false. Their predictions may be true in some contexts and false in others, so the goal of political scientists is to determine what each model is useful *for*. In my view, models that make false predictions about phenomena they intended to explain, or that make predictions that are usually false, are themselves false.

References

Bennett, Andrew and Bear F. Braumoeller. 2009. "Where the Model Frequently Meets the Road: Combining Statistical, Formal, and Case Study Methods." Washington, DC: Georgetown University.

Clarke, Kevin A. and David M. Primo. 2007. "Modernizing Political Science: A Model-Based Approach." *Perspectives on Politics* 5:4, 741–753.

España-Nájera, Annabella, Xavier Márquez, and Paul Vasquez. 2003. "Surveying the Field: Basic Graduate Training in Comparative Politics." *APSA-CP: Newsletter of the APSA Organized Section in Comparative Politics* 14:1, 28–34.

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

Green, Donald P. and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven: Yale University Press.

Lakatos, Irme. 1970. "Falsification and the Methodology of Scientific Research Programmes." In *Criticism and the Growth of Knowledge*. Imre Lakatos and Alan Musgrave, eds. (Cambridge: Cambridge University Press), 91–196.

Lieberman, Evan. 2005. "Nested Analysis as a Mixed-Method Strategy for Cross-National Research." *American Political Science Review* 99:3, 435–452.

Mahoney, James. 2003. "Knowledge Accumulation in Comparative Historical Research: The Case of Democracy and Authoritarianism." In *Comparative Historical Analysis in the Social Sciences*. James Mahoney and Dietrich Rueschemeyer, eds. (Cambridge: Cambridge University Press), 131–174.

Does It, Really? Measurement Error and Omitted Variables in Multi-Method Research

David Kuehn
University of Heidelberg, Germany
kuehn@uni-heidelberg.de

Ingo Rohlfing
University of Cologne, Germany
rohlfing@wiso.uni-koeln.de

A recurring tenet in much of the recent discussion on combining case studies and regression analyses and, to a lesser degree, case studies and QCA, highlights the purported ability of Multi-Method Research (MMR) to overcome the classic problems attached to each of the individual methods. The large-N part is employed to uncover cross-case regularities and causal effects, while in-depth case studies are undertaken to identify causal mechanisms in a subset of cases. In this, one purpose of case studies is to cross-validate the insights of the large-N analysis by identifying possible measurement error and exposing potentially omitted variables. Thereby, it is assumed, case studies are able to make the large-N analysis more robust.

In this essay, we critically revisit these arguments from a methodological perspective. We argue that the ability of case studies to effectively enhance the inferential quality of the large-N method is significantly limited due to the very problems that they are supposed to solve: measurement error and omitted variables. While there is some discussion of these two issues on the cross-case level (King, Keohane, and Verba 1994; Lieberson 1991), there is only limited and largely implicit reflection on their relevance for the within-case level. This is unfortunate, as measurement errors and omitted variables on the within-case level are no less damaging for the inferential power of case studies as they are for large-N analyses. Measurement error and omitted variables are particularly likely to occur in process tracing because of problems in the use of sources for data collection and the generation of inferences on the basis of this data. We, therefore, maintain that the ability of the case study to serve as a double check on the quantitative part of MMR designs is overestimated. To be sure, many of the issues we highlight in this paper have been thrown up in the methods debate before. However, we think there is an insufficient transfer of the arguments made in the realm of the case study literature to the work on MMR, which is what we aim to achieve in the following.

This essay proceeds in four parts. In the first section, we shortly revisit the methodological literature to extract the two main functions of case studies as robustness checks in MMR designs: the identification of measurement error and the uncovering of omitted variables. In parts two and three, we discuss to what extent process tracing can deliver what the MMR literature promises. The fourth section concludes with a short summary of the argument.

The Role of Case Studies in MMR

Given that there is a variety of possible methodological approaches which can be integrated into a single research design (cf. Johnson, Onwuegbuzie, and Turner 2007), most of the recent discussion of MMR—our essay included—is concerned with the combination of statistical techniques, and less often QCA, and in-depth case studies (e.g. Bennett 2002; Lieberman 2005; Schneider and Rohlfing 2009). In these designs, the case study part serves two related purposes. Its primary function is to establish a causal relationship by elucidating causal mechanisms. This role addresses the well-known caveat that a cross-case regularity does not necessarily reflect causation and mirrors the increasing consensus on the relevance of mechanisms for making causal claims. Observed cross-case regularities do not provide a solid foundation for inferring causality unless it can be empirically demonstrated that the purported cause is linked to the outcome via an uninterrupted process (George and Bennett 2005; Hall 2006). Case studies are claimed to provide this foundation by tracing the processes through which the causes produce the observed effects. The second function of case studies is to cross-validate the specification of the cross-case model by checking for measurement error and omitted variables, thereby improving the robustness of the large-N results (Fearon and Laitin 2008; King, Keohane, and Verba 1994: 152–183). In this essay, we focus on the latter issues of measurement error and omitted variables on the within-case level. The issue of identifying causal mechanisms through case studies is dealt with more extensively in another paper of ours (Kuehn and Rohlfing 2009).[1]

Measurement error is understood as the use of indicators which are either not reliable or lack validity in capturing the empirical content of the underlying concept (Brady and Collier 2004: 295). Validity problems may derive from mistakes in concept formation, for example when important attributes and, thus, indicators are missing or redundant, or from flawed measurement, for example if an indicator with low content validity is used (Adcock and Collier 2001). The dangers of measurement error for the robustness of statistical findings are well documented. Depending on whether measurement error is systematic or non-systematic, whether it occurs on the dependent or the independent variable, and whether it is correlated across variables, the estimates can be rendered inefficient, biased, and inconsistent (Herrera and Kapur 2007). QCA suffers from severe problems as well when measurement error is present (Seawright 2005). To some extent, the problem of measurement error is inherent to social science research simply because the concepts used to systematize the empirical world are constructs made by researchers. Accordingly, there will always be some degree of imprecision involved in fitting theoretical constructs to the empirical phenomena of interest (King, Keohane, and Verba 1994: 152).

In large-N research, there are statistical means by which measurement error can be tested and corrected for (Rabinovich 2000).[2] The MMR literature, however, suggests in-depth analysis of one or a few number of cases as an additional check for the quality of concepts and the validity of indicators (Coppedge 1999; McKeown 1999). It is to note that the reliability of an indicator cannot be evaluated through case studies as this would require repeated measurement of the same concept in more than one or a few cases. As regards concept formation, it is claimed that case studies can provide valuable information on construct validity. This can be done, for instance, by paying attention to context and thus avoiding conceptual stretching (Adcock and Collier 2001). Concerning the scoring of indicators, process tracing is held as less prone to measurement error, "because it can intensively assess a few variables along several qualitative dimensions, rather than having to quantify variables across many cases" (George and Bennett 2005: 220). In sum, the ability of case studies to provide superior concepts and measurement has been identified as one of the most important contributions that method has to offer (King, Keohane, and Verba 1994: 151–168).

Similar arguments are made for the problem of omitted (or missing) variables, which undermines the cross-case analysis in QCA and produces biased estimates in regression analysis (Seawright 2005).[3] As is the case for measurement error, there are statistical tools to check for the existence of omitted variables (Clarke 2005), but case studies are argued to be a particular suitable tool to uncover hitherto neglected factors in the cross-case analysis (George and Bennett 2005: 34). Furthermore, as the range of possible confounders is in principle unlimited (King, Keohane, and Verba 1994: 174), case studies are supposed to identify exactly *which* variables should be included into the regression model. In this, both typical and deviant cases are deemed to be particularly useful for the search of omitted variables. Process tracing in typical cases is undertaken to check for spurious cross-case patterns, that is, the working of a third variable which influences both the independent and dependent variable. The analysis of deviant cases can help to identify variables which can then be included into the model in order to enhance model fit (Lieberman 2005).

Measurement Error in Within-Case Analysis

Thus far, the literature discussed the implications of measurement error with respect to cross-case inferences only (King, Keohane, and Verba 1994; Lieberson 1991). Measurement errors are usually attributed to the researcher's "slip of the hand," that is, random mistakes in coding the cases, or to the use of weak concepts or invalid indicators. Besides these general arguments there are, to our knowledge, no guidelines of how actually to perform a within-case analysis in order to improve concepts and measurement. We argue that the rather optimistic view on the potential of process tracing is not warranted for two reasons. First, the case study suffers from a variant of the classic small-N problem because the quality of a concept and measurement must be assessed in a larger set of cases. Second, within-case analysis is similarly prone to measurement error as large-N analysis, which undermines its utility for cross-validation.

Concerning the first problem, process tracing may in fact show that a concept is misspecified or that a weak indicator has been used. Intimate knowledge of cases provides the ground for potentially more realistic interpretations and more valid classifications of empirical facts, for example by identifying functional equivalents which would go undetected in the standardized coding procedures typical for large-N studies (Adcock and Collier 2001). A comparative case study interested in political opposition, for instance, may suggest that in some countries this concept is better measured through the behavior of parties and that in other countries interest group politics might be a more appropriate indicator (Verba 1967). Ultimately, however, the validity of the indicators must be evaluated by calculating their convergent or discriminant validity or, more preferably, construct validity (Adcock and Collier 2001). These tests are indispensable because concepts or indicators are not supposed to fit one or a small set of cases particularly well. Instead, conceptualization and measurement should be as widely applicable as possible to the *population* of interest (Carmines and Zeller 1979). If, for example, a given concept does not fit some individual cases but is valid for a large share of a given population, it would be of little use to reconceptualize it in order to make it fit the deviances. If at all, it would be considered more plausible to eliminate these cases on theoretical and conceptual grounds. At the same time, however, the latter strategy runs the danger to ad hoc delineation of the population. There probably are always cases for which the concept or indicators are not optimal, but the lack of fit does not warrant their exclusion.

Process tracing employed for improving concepts and measurement, therefore, is confronted with the same small-N problem known from causal inference (cf. Lieberson 1991): even if one finds that the concept does not fit well one or a few numbers of cases it is not clear which conclusions should be drawn without knowledge about the generalizability of what one learned from the cases under scrutiny. Of course, qualitative within-case analysis is particularly valuable both for causal inferences and the improvement of concepts and measures with respect to the cases under analysis, that is, *internal validity*. However, the degree to which these insights are externally valid is impossible to gauge with much certainty because such an evaluation would include statements about cases which have not been subject to empirical analysis.

The second neglected problem of case studies is measurement error on the within-case level. Measurement means assigning scores to indicators, which is not necessarily as unproblematic in small-N research as some MMR researchers seem to believe. Measurement error in case studies can derive from two sources: the deficient collection of empirical data and the incorrect interpretation of this data. Concerning the first problem, it is well known that inferences drawn from process tracing are necessarily only as good as the sources on which they are based. The specific problems typical for different sources of empirical evidence, e.g. primary and secondary sources, newspapers, and interviews, have been extensively discussed elsewhere and need not be elaborated upon here (e.g., Lustick 1996; Thies 2002). The usual recommendation to cope with these problems is triangulation, that is, to draw on multiple sources, thereby addressing their respective shortcomings. Case studies are supposed to be particularly suited to triangulate data as they offer "an opportunity to fact check, to consult multiple sources, to go back to primary materials, and to overcome whatever biases may affect the secondary literature" (Gerring 2007: 59). Triangulation certainly may help to ameliorate the problems of data collection in process tracing, but a certain, and often considerable degree of uncertainty about the full picture remains. It is exactly this uncertainty from which measurement error in small-N research derives. For illustrative purposes, assume that one is interested in the prevalence and severity of human rights violations in order to code the democratic quality of a country. Due to the very nature of many autocratic regimes, it will be difficult to obtain sufficient data to effectively measure the degree of human rights violations in those countries as official documents will be hard to obtain, newspapers and media reports will likely be subject to censorship, and interviewees will often be reluctant to share information with foreign researchers for fear of repression and retribution. In fact, the problems related to validly measuring democratic quality are systematically related to the concept itself. In this example, it is likely that the available empirical evidence biases the classification of the regime, possibly suggesting a better democratic quality than the actual state of civil rights would actually allow. The difference between the observed and the true state of democratic quality in that country, however, cannot be validated based on the available empirical evidence and thus it is not possible to eliminate the existing measurement error. In other examples, the limited availability and inherent bias of sources might not be as evident as in this example and measurement error might occur without the researcher being aware of it.

But even if sufficient empirical evidence is accessible to the researcher, measurement error on the within-case level might just as well derive from subsuming a certain observation under a wrong concept or interpreting some piece of empirical evidence wrongly. This danger might be at least as prevalent in process tracing as it is for cross-case research. First, empirical observations on the within-case level are usually highly disparate in nature and are not typically easily compared with each other. Second, and in contrast to observations on the cross-case level, the rules of how to identify, interpret and code within-case observations are usually not formalized, standardized and made explicit (Collier, Brady, and Seawright 2004; King and Powell 2008). Therefore, the replication of measurement in order to uncover measurement errors in within-case analyses is often hardly possible.

As an example, Grieco (1990) performs a deep and rich case study of how some agreements of the GATT Tokyo Round were implemented in the 1980s. He derives multiple observable implications from two competing theories, neo-realism and neoliberalism, and claims that the empirical evidence is more in accord with the neo-realist explanation. However, while the empirical within-case evidence he provides supports the neorealist argument, it is not conclusive as the process observations he cites can equally well be subsumed under the plausible competing hypothesis that trade negotiations are actually shaped by domestic politics and the interests of economic actors. For instance, Grieco quotes a report of an EU commission saying that "the 'cost of non-Europe' represents a direct, serious threat to essential components of the Community's industrial potential." Without knowing whether the "Community's industrial potential" is deemed important for the sheer survival of Europe, which is what neo-realism predicts, or whether this statement simply reflects the influence of strong lobby groups aiming to operate in a large, liberalized market, this piece of evidence is inconclusive and subsuming it under one of the competing accounts runs the risk of committing a measurement error. The problem of misinterpretation of process evidence is particularly salient for those within-case hypotheses which "bottom out" (Machamer, Darden, and Craver 2000) on the level of individual or collective actors and try to explain social phenomena by these actors' intentions. Njølstad (1990), for instance, shows that such problems are ubiquitous even in a well-researched area like US nuclear policy which is replete with disagreements on how to interpret the supposedly well-document interests and beliefs of even the most important actors and how these may have contributed to even the most critical and obvious events and developments.

Omitted Variables in Within-Case Analysis

Similar rationales hold for the problem of omitting variables. It might be that the researcher is unable to identify a relevant factor due to sheer lack of information in the available sources. Since the researcher usually is not able to decide if she has collected a sufficient amount of evidence, the existence of one or more highly important pieces of evidence may go undetected (Thies 2002). Based on limited within-case evidence, a researcher could fall for a type I error by inferring a causal process between X and Y, while in fact the evidence is flawed or there is some hidden evidence that the actors concealed their true motivations. At the same time, there is a risk of a type II error by erroneously accepting the null hypothesis when indeed there was some causal process which remained undetected because of unavailable evidence.

Either way, the absence or presence of seemingly convincing within-case evidence does not necessarily mean that the identified causal process (or its absence) is true. Assume, for instance, that one wants to test two competing explanations of the democratic peace thesis on the within-case level. If the empirical evidence is compelling for one account and weak for the other, one would infer that process tracing corroborates the first explanation and disconfirms the second one. However, if these inferences and scores are primarily grounded in the accessibility of evidence in favor of the first explanation, one would draw erroneous conclusions. This could be, for example, the case when one explanation focuses on civil society activities in democracies and the other hypothesizes peaceful effects of economic ties between democratic countries. Propeace protests of civil society would be quite visible because their actions take place in public. At the same time, however, there may be massive lobbying behind the scenes by economic interest groups that fear to lose from war. Furthermore, from the perspective of politicians, it might be more attractive to launch a peace initiative by referring to civil society and democratic norms instead of economic interests and pressure group lobbying. In this example, process tracing is more likely to uncover empirical evidence in favor of the civil society explanation and an important variable would be omitted from the theoretical account. However, if politicians and economic actors in question conceal their true motivations in interviews, and if primary sources are kept secret, one cannot do more than speculate about the actual role of economic interests.

Conclusion

In much of the recent MMR literature, case studies are promoted as adequate tools to check the regression results for measurement error and omitted variables. We have argued that the method might not be as capable to fulfill this function as it suffers from similar weaknesses as large-N methods when it comes to handling inferential challenges. As a particularly pressing problem we have identified the quality and use of sources. They affect the possible contribution case studies can make to overcome measurement errors and to identify omitted variables. Therefore, the opportunity of within-case analyses to cross-validate the large-N part is limited. This does not mean that qualitative within-case analyses cannot improve one's trust in the large-N findings at all and that it therefore should not be undertaken for these purposes. However, we want to caution against putting too much hopes in the insights gathered from process tracing and recommend to additionally rely on the established diagnostic tests for measurement, omitted variables, and other specification issues in regression analysis (Rohlfing 2008).

Notes

[1] By focusing on the second function of case studies we are, however, able to make inferences about the first function because they are closely related. If we can show that the case study part in MMR designs is not able to test the robustness of the large-N analysis, the method will neither be able to fulfill the even more demanding task of founding causal processes.

[2] No such means are available for QCA at present.

[3] Provided that the omitted variable is correlated with the included variables, which is almost always the case, the consequences of omission are ambiguous. It is only in the bivariate case that the estimate is biased toward zero (King, Keohane, and Verba 1994: 171).

References

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

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

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

Collier, David, Henry E. Brady, and Jason Seawright. 2004. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Henry E. Brady and David Collier, eds. (Lanham: Rowman & Littlefield), 229–266.

Carmines, Edward G. and Richard A. Zeller. 1979. *Reliability and Validity Assessment*. Beverly Hills, CA: Sage Publications.

Clarke, Kevin A. 2005. "The Phantom Menace: Omitted Variable Bias in Econometric Research." *Conflict Management & Peace Science* 22:4, 341–352.

Coppedge, Michael. 1999. "Thickening Thin Concepts and Theories: Combining Large N and Small in Comparative Politics. *Comparative Politics* 31:4, 465–476.

Fearon, James D. and David Laitin. 2008. Integrating Qualitative and Quantitative Methods. In *The Oxford Handbook of Political Methodology*. Janet M. Box-Steffensmeier, Henry Brady and David Collier, eds. (Oxford: Oxford University Press), 756–778.

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

Gerring, John. 2007. *The Case Study Method: Principles and Practices*. Cambridge: Cambridge University Press.

Grieco, Joseph M. 1990. *Cooperation Among Nations*. Ithaca: Cornell University Press.

Hall, Peter A. 2006. "Systematic Process Analysis: When and How to Use It." *European Management Review* 3, 24–31.

Herrera, Yoshiko M. and Devesh Kapur. 2007. "Improving Data Quality: Actors, Incentives, and Capabilities." *Political Analysis* 15:4, 365–386.

Johnson, R. Burke, Anthony J. Onwuegbuzie, and Lisa A. Turner. 2007. "Toward a Definition of Mixed Methods Research." *Journal of Mixed Methods Research* 1:2, 112–133.

King, Gary and Eleanor Neff Powell. 2008. "How Not to Lie Without Statistics." Paper presented at the Annual Meeting of the American Political Science Association.

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

Kuehn, David and Ingo Rohlfing. 2009. "Causal Explanation and Multi-Method Research in the Social Sciences." Paper presented at the Annual Meeting of the American Political Science Association.

Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99:3, 435–452.

Lieberson, Stanley. 1991. "Small Ns and Big Conclusions: An Exami nation of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70:2, 307–320.

Lustick, Ian. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias." *American Political Science Review* 90:3, 605–618.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67:1, 1–25.

McKeown, Timothy J. 1999. "Case Studies and the Statistical World View." *International Organization* 53:1, 161–190.

Njølstad, Olav. 1990. "Learning from History? Case Studies and Limits to Theory-Building." In *Arms Races: Technological and Political Dynamics*. Nils Petter Gleditsch and Olav Njølstad, eds. (Oslo: Sage), 220–246.

Rabinovich, Semyon G. 2000. *Measurement Errors and Uncertainties: Theory and Practice*. 2nd ed. New York: AIP Press.

Rohlfing, Ingo. 2008. "What You See and What You Get: Pitfalls and Principles of Nested Analysis in Comparative Research." *Comparative Political Studies* 41:11, 1492–1514.

Schneider, Carsten Q. and Ingo Rohlfing. 2009. "Explaining Party Competition: A Multi-Method Approach Combining Fuzzy Set Qualitative Comparative Analysis (fs/QCA) and Case Studies." Paper presented at Conference of the European Consortium for Political Research.

Seawright, Jason. 2005. "Qualitative Comparative Analysis vis-à-vis Regression." *Studies in Comparative International Development* 40:1, 3–26.

Thies, Cameron G. 2002. "A Pragmatic Guide to Qualitative Historical Analysis in the Study of International Relations." *International Studies Perspectives* 3:4, 351–372.

Verba, Sidney. 1967. "Some Dilemmas in Comparative Research." *World Politics* 20:1, 111–127.

Two Cultures: Concepts, Semantics, and Variable Transformations

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

This is the first in a series of essays that will appear in the QMMR section newsletter.[1] They will typically continue the analysis Jim Mahoney and I started in our "Two Cultures: Contrasting Quantitative and Qualitative Research" article (2006). Sometimes the essay will contrast the two cultures; sometimes it will discuss practices that are unproblematic in one culture, but which appear contestable in the other. I request the indulgence of many members of the section; the view of "qualitative culture" presented here certainly does not speak for all members of the section. There is a wide diversity of views about what constitutes qualitative methods, those expressed here ultimately are my own about the practices of both qualitative and quantitative methods.

This essay looks at practices of variable transformations such as standardization and logging, which are virtually never problematic in statistical research. Not surprisingly, given my interests, I relate these practices to the concepts that the transformed variables are supposed to represent or measure. I ask whether the transformed variables do a good job of capturing the meaning of the underlying concepts of interest.

We have customs and norms (in the anthropological sense) within statistical culture about transforming variables that make practices such as logging and standardization common and noncontroversial. When these practices are viewed from a qualitative perspective with its emphasis on concepts, semantics, and meaning, however, these practices become problematic. Variable transformations in the quantitative culture respond to imperatives of statistics; qualitative scholars work under a different set of interests that focus on concepts, definitions, and the meaning embodied in variables. This culture provides a different interpretation of what it means to, say, log common variables such as GDP/capita, or standardize the polity democracy measure. These different views on variable transformations support Mahoney and my two-cultures hypothesis.

In making this contrast, I follow Charles Ragin's (2000, 2008) suggestion that fuzzy logic provides a natural set of tools for getting from raw data, such as GDP/capita, to numeric values for concepts, such as national wealth or development. Fuzzy logic at its origin (Zadeh 1965; see Kosko (1993) or McNeill and Freiberger (1994) for very accessible introductions) was created in part as a mathematical theory of semantics. As such it is an appropriate tool for transforming raw data and indicators into numbers that better match the theory and meaning of key theoretical concepts.

In this essay I introduce the Fundamental Principle of Variable Transformation. This principle requires that all transformations of variables be meaning preserving or increasing. To use one of my examples below, the principle requires that if one logs GDP/capita it should better represent what the scholar means by, say, wealth than the untransformed data. From the qualitative perspective all transformations which do not conform to this principle are suspect.

Meaning Transformations versus Statistical Transformations

Befitting the two-cultures metaphor, the language used to talk about the "same problem" varies significantly across the traditions. Within the quantitative culture one speaks of "variables" (aka concepts) and "indicators" (aka numeric data). Typically, the nature of the relationship between variables and and indicators is not discussed. For example, is the relationship between the indicator and variable causal? Is it merely correlational? In practice, one often laments that the indicator does not reflect, measure, or represent well the theoretical variable.

Within the qualitative culture one would ask about the semantic, theoretical relationship between the numeric data, say GDP/capita, and a concept such as economic development. What is the definition of economic development, and which attributes of this concept are measured by GDP/capita? How do different levels of GDP/capita relate to the concept of economic development?

Thus, both qualitative and quantitative cultures can take the same data, e.g., GDP/capita, as somehow a measure or related to wealth, but they then proceed to do quite different things. Within the statistical culture GDP/capita as a measure of wealth is commonly transformed, e.g., logged, for statistical reasons. A qualitative scholar will take the same data and and ask about the theoretical and semantic relationship between GDP/capita and the concept of wealth. Fuzzy logic is then a means to transform, or calibrate in Ragin's terminology, these raw data to better reflect what the theorist means by wealth.

Standardization and Concept Meaning

Perhaps one of the most popular transformations is to standardize a variable. In some fields it is quite common, if not virtually obligatory, to standardize variables before performing statistical analysis. Standardization often does not change the statistical results, because most parameter estimates retain their statistical properties such as unbiasedness when the variable is subject to a linear transformation.[2] To choose an example, perhaps familiar to members of the QMMR section of APSA, I take Gerring's (2005, chapter 5) advice to standardize variables in order to select case studies based on their "extreme" values. Gerring uses the concept of democracy as coded and measured by the polity project (Jaggers and Gurr 1995).

Let us recall what is going on when a variable is standardized. The formula is $(x_i - \bar{x}) / s$, where $\bar{x}$ is the mean and s is the standard deviation. With the Polity scale, for example, all cases are coded from –10 to 10 according to their level of democracy. Standardization converts these numbers into a scale of standard deviations from the mean.

It is critical to note that standardization is based on data, not on semantics. The mean, as well as the standard deviation, depends on the data: different datasets can vary significantly on these values. So in one dataset observation a standardized x_i could have a different value than it does in another dataset, e.g., in one it could have value –1, while in another +1. From a qualitative or semantic perspective this can seem odd, since why should the democracy level of a country depend on how other countries happen to be coded. Likewise, a country might not experience any change, yet its democracy coding can change because other countries' democracy levels are changing. To see this in practice, take the semantic meaning embedded in the Polity rule that "full democracy" characterizes cases with values from 7–10. Using standardized values involves saying something like democracy is those cases, say, 1.65 standard deviations or more from the mean.

Gerring's case study goal is to choose *extreme* cases. Standardization is an obvious choice because we have some ideas about what extreme means in terms of standard deviations: an observation that is 2–3 standard deviations from the mean is extreme. This of course depends on the sample mean, which is considered to be the "middle point." The sample mean may, or may not, correspond to the conceptual or theoretical middle point. In Gerring's concrete example, the sample mean is about +3 on the polity scale of –10 to 10, while the conceptual middle point might be zero.

One consequence of Gerring's standardization proposal is that the most extreme cases are always authoritarian regimes. Because the sample mean is about +3, the authoritarian cases of –10 always have larger standardized values than do the complete democracies cases with value 10. From the qualitative perspective standardization is counterproductive; the extreme values are "obviously" –10 and 10 for authoritarianism and democracy respectively.

In short, the qualitative, semantic approach will generally see standardization as a step backwards. It introduces irrelevant considerations, i.e., the potentially changing distribution of the data, into what is a semantic relationship. Standardizing data for the selection of extreme cases in this way violates the Fundamental Principle of Variable Transformation. It brings into semantic and theoretical relationships aspects of the real world distribution of data that are irrelevant and potentially misleading.

Logging Variables: Is Clear-Cutting Acceptable?

Logging a variable is very common, and often recommended in statistical research. To take my main example, the decision to log GDP/capita is rarely controversial. Although not all scholars log it, one can find countless examples of research of all kinds where scholars do.

On the qualitative side, fuzzy logic provides a basis for asking about the semantic or theoretical relationship between GDP/capita and the concept of wealth.[3] While it is beyond the scope of this essay to give a good introduction, fuzzy logic asks about the relationship between given numeric scales or data and the theoretical or semantic concept. For example, what is the relationship between GDP/capita data and the concept of wealth? By convention zero means that a country has no membership in the set of wealthy countries, while one means full membership (see Ragin 2000, 2008 for an introduction, or any textbook on fuzzy logic). Hence we can use information like GDP/capita to assess the extent to which a country is a full or partial member of the set, "wealthy country." This usually involves transforming the raw data, or what Ragin (2008) calls calibration.

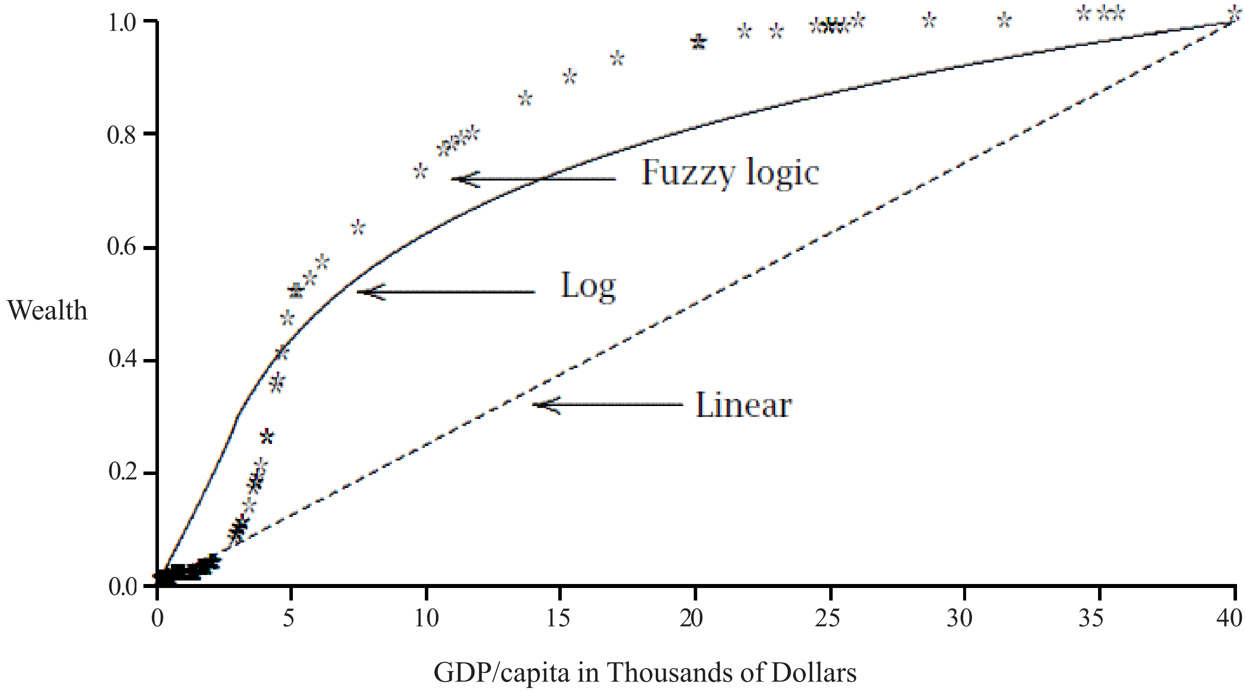
From the fuzzy logic point of view, the quantitative culture automatically assumes a linear, semantic relationship between GDP/capita data and the concept of wealth, as illustrated in figure 1 by the dashed line. The usually unstated assumption is that wealth increases linearly with GDP/capita. If one logs the data then one potential interpretation (see below for another) is that there is a decreasing returns relationship between GDP/capita and wealth as illustrated by the solid line in figure 1. A very popular fuzzy logic relationship is the S-shaped curve between GDP/capita and wealth (taken from Ragin 2008, chapter 5).

The fuzzy logic approach does not assume a linear relationship between GDP/capita and wealth. Its approach to coding the data is more similar to the diminishing returns argument that could be used for logging. The fuzzy logic intuition is that the GDP/capita difference between Switzerland and Germany is not important; they are both 100 percent wealthy countries. Hence variation among cases at the upper end is irrelevant.

Logging has this sort of relationship at the high end but not the low end. Here is a fundamental difference with fuzzy logic. In the fuzzy logic view of wealth, differences at the low end have little meaning, since all of these countries are clearly not wealthy. For example, the difference in wealth between Chad and Mali is irrelevant; they both have zero membership in the category, "wealthy country." But for the scholar who logs GDP/capita, small differences at the lower end are accentuated and made larger by the logging transformation (as indicated by the steep slope for the log curve in figure 1). One view is not inherently right; but the fuzzy logic view calls attention to these differences and forces scholars to justify their decisions. As figure 2 below shows the vast majority of the data are in the low end, so fuzzy logic scores on wealth will usually be quite different from the linear or log ones. Here we see another difference between the two cultures: the quantitative school implicitly assumes a linear or log relationship while the qualitative scholar is much more likely to use some other nonlinear relationship, such as the S-curve in figure 1.

It is important to note that the S-curve, indicated by the stars in figure 1, is not symmetric. The concept of wealth is not the negation of the concept of poverty. Poverty is a separate concept which would have its own relationship with GDP/capita. This is another big difference between the two traditions. In a typical study of poverty, for example, an economist

Figure 1: Variable Transformations: A Fuzzy Logic Perspective on GDP/capita and Wealth



Source: Fuzzy logic curve from Ragin (2008: chapter 5).

would use the same membership function as she would in a study of wealth, i.e., she would use GDP/capita for both. By contrast, a fuzzy logic analyst would almost never do that since poverty and wealth are two separate concepts; one is not the simple mirror or inverse of the other.

So how does one go about determining the semantic relationship between raw data like GDP/capita and wealth? For example, how does one decide where to draw the cut-off points for cases with full membership in a category? Ragin (2008) provides an example. He explores how scholars and important institutions like the World Bank determine if countries are rich or poor. He uses their semantic practices to help determine the shape of the S-curve in figure 1. Looking at the practices of scholars and major institutions thus informs Ragin's semantic transformation. He uses that information to make the new variable more faithful to the way researchers and institutions use GDP/capita to conceptualize wealth or economic development.

Thus the fuzzy logic approach to data transformation satisfies the Fundamental Principle of Variable Transformation. One transforms raw data so that they match better what the analyst means by the concept. It is possible to preserve or enhance the meaning of the underlying concept with a linear, log, or standardization transformations, but the choice needs to be justified.

Above I argued that for a qualitative scholar logging is about the relationship between data and concept semantics. This question does not exist really in quantitative culture (it could but it would be rare). So what are the interpretation and uses of logging if they do not involve conceptualization and semantics? How do quantitative scholars justify the practice of logging?

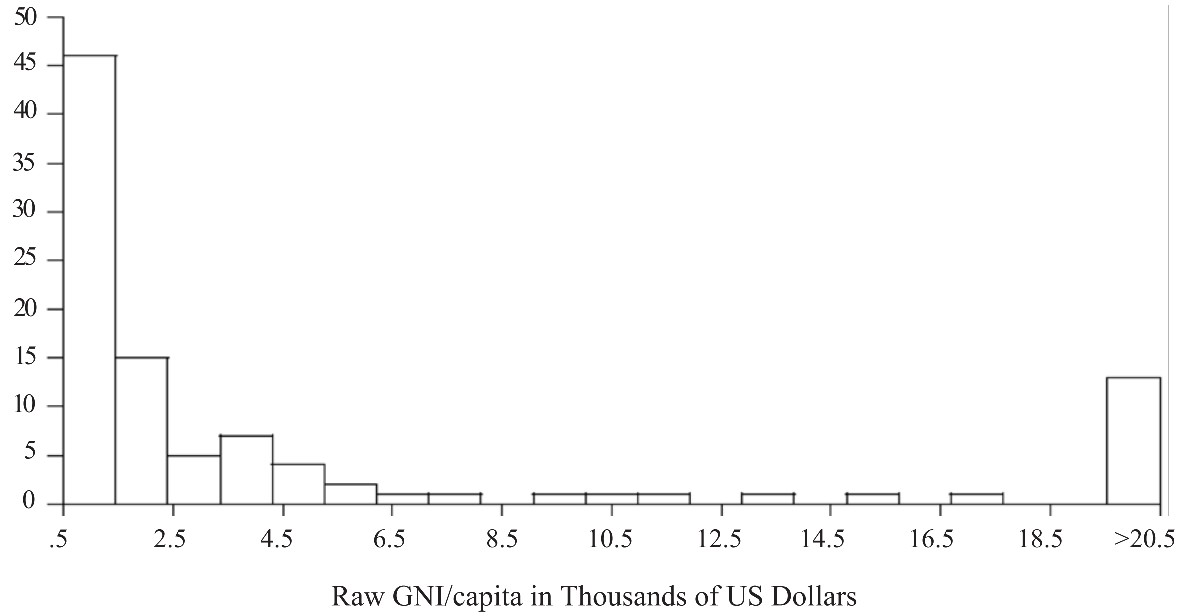
Sometimes logging arises very clearly for theoretical reasons. For example, Jones et al. (2009) argue, and empirically show, that almost all government budgets follow a power law distribution. This distribution – $y \sim ax^b$ – means that the obvious empirical test would be to log both sides of the equation. This is clearly a reasonable practice.

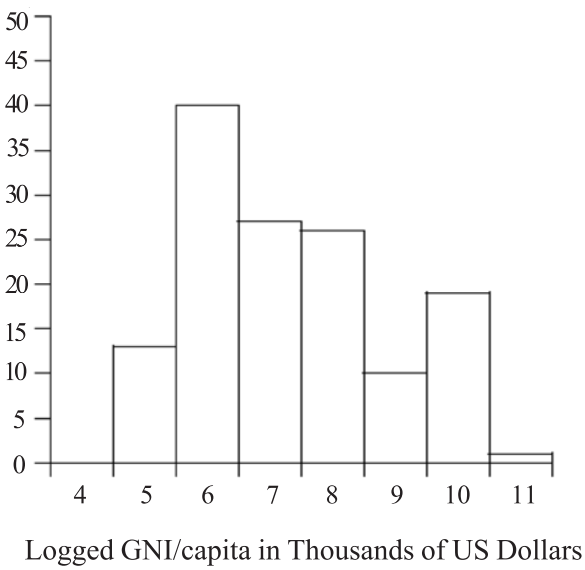
Another common rationale for logging is that the data are skewed, as is true of the GDP/capita data. Unlike the Jones example, however, this rationale is based on the distribution of the data, not a theoretical reason. It seems very similar to the standardization example discussed above: one is transforming the scale for reasons that have nothing to do with the semantic relationship between the raw data and the concept, nor for theoretical reasons as in the power law example. So, inherently the qualitative school is tempted to see this use of logging as violating the Fundamental Principle of Variable Transformation.

However, if your research goal involves using certain statistical methods then skewness of the data can be an issue. So within this research context and culture logging makes perfect sense because it does a good job at removing skewness, as illustrated in figure 2. Figure 2 takes Ragin's data that are used to give the starred values in figure 1. Hence fuzzy logic and statistics transform data, but how they do it and the rationales for the transformations differ radically.

Related to skewness is the rationale that logged data fit "better" in statistical analyses, where better means stronger substantive impact or higher significance levels. This sort of rationale falls under what economists call specification searches, which have been extensively debated over the decades. One could imagine doing all sorts of variable transfor-

Figure 2: Logging as a Correction of Skewness: GNI/capita





mations and then picking the one that gives the strongest results. The practice of logging to improve fit seems to generate little objection. For example, Kurtz and Schrank (2007) argue that governance is not related to economic growth. The World Bank economists in their response say, "In the next panel we show the effect of two minor departures from the original KS specification. Instead of entering per capita GDP in levels as they do, we enter it in log-levels. This is very standard practice in cross-country empirics and statistically is more appropriate since the relationship between the dependent variable and log per capita GDP is much closer to being linear, and we are using a linear regression model" (Kaufman, Kraay, and Mastruzzi 2007: 59).

From the qualitative, semantic perspective there are two questions that the statistical culture fuses into one. The first in logical order is the relationship between GDP/capita data and the concept of wealth; the next question is the relationship between wealth and the dependent variable. Because the first question does not exist in the quantitative culture, the interpretation of logging almost exclusively involves the second question. The qualitative scholar says that one needs to fix upon the relationship between GDP/capita and wealth before moving to the relationship between wealth and a dependent variable.

I hope this essay incites all to ask about the rationales for variable transformations. In particular, it is useful to verify that the transformation conforms to the Fundamental Principle of Variable Transformation. Does the transformation increase the validity of the resultant numbers? Does it make the new calibrated or transformed numbers closer in meaning to the theoretical concept?

Notes

[1] Thanks to Jan Box-Steffensmeier, Jim Mahoney, Charles Ragin, and Chad Westerland for comments on earlier versions.

[2] However, if the variable is further subject to transformation or analysis this may no longer the case e.g., if the transformed variable is used in interaction terms.

[3] GDP/capita is used as indicator of a wide variety of concepts, but I stick to one of the closest ones.

References and Suggested Readings

Brown, J. David, John S. Earle, and Scott Gehlbach. 2009. "Helping Hand or Grabbing Hand? State Bureaucracy and Privatization Effectiveness." *American Political Science Review* 103:2, 264–283. Here is an example where almost all the variables are logged.

Busse, Matthias and Carsten Hefeker. 2007. "Political Risk, Institutions and Foreign Direct Investment." *European Journal of Political Economy* 23:2, 397–415. They want to log the foreign direct investment variable, but that can take on negative values so they transform the data using $y = \log(x + (x^2 + 1)^{.5})$.

Carment, David, Yiagadeesen Samy, and Stewart Prest. 2009. "State Fragility and Implications for Aid Allocation: An Empirical Analysis." *Conflict Management and Peace Science* 25:4, 349–373. See how logged and standardized infant mortality is used as an indicator for standard of living.

Fearon, James. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth* 8:2, 195–222. He argues that one measure of ethnic diversity is better than another because the distribution of cases is not skewed.

Fischer, Claude S., et al. 1996. *Inequality by Design: Cracking the Bell Curve Myth*. Princeton: Princeton University Press. Critiques the normalization of the Bell Curve data.

Gerring, John. 2005. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.

Herrnstein, Richard and Charles Murray. 1994. *Bell Curve*. New York: Free Press. The intelligence test scores they use are not normally distributed, so they transform the scores so that they are normally distributed, see Fischer et al. (1996) for a critique.

Jaggers, Keith and Ted Robert Gurr. 1995. "Tracking Democracy's Third Wave with the Polity III Data." *Journal of Peace Research* 32:4, 469–482.

Jones, Bryan D., et al. 2009. "A General Empirical Law of Public Budgets: A Comparative Analysis." *American Journal of Political Science* 53:4, 855–873.

Kaufman, Daniel, Aart Kraay, and Massimo Mastruzzi. 2007. "Growth and Governance: A Reply." *Journal of Politics* 69:2, 555–562.

Kosko, Bart. 1993. *Fuzzy Thinking: The New Science of Fuzzy Logic*. New York: Hyperion.

Kurtz, Marcus J. and Andrew Schrank. 2007. "Growth and Governance: Models, Measures, and Mechanisms." *Journal of Politics* 69:2, 538–554.

Londregan, John and Keith Poole. 1996. "Does High Income Promote Democracy?" *World Politics* 49:1, 1–31. To be able to use their preferred statistical methods they transform the polity scale using the following formula, where S is the polity democracy score: $T(S) = \log(S + 10.5) - \log(10.5 - S)$.

Mahoney, James and Gary Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14:3, 227–249.

McNeill, Dan and Paul Freiberger. 1994. *Fuzzy Logic*. New York: Simon and Schuster.

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

Ragin, Charles. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press. See chapter 5 for an extensive discussion of fuzzy logic variable transformations and calibrations.

Zadeh, Lotfi A. 1965. "Fuzzy Sets." *Information and Control* 8:3, 338–353.

Area Studies, Comparative Politics, and the Role of Cross-Regional Small-N Comparison

Rudra Sil
University of Pennsylvania
rudysil@sas.upenn.edu

Stephen Hanson's (2008) essay in the Fall 2008 issue of this newsletter offers a much needed corrective for the perception that area-focused scholarship is neither rigorous enough nor theoretical enough to deserve a prominent place in the discipline of political science. While concurring with Hanson's eloquent defense of the value of area studies, this essay is motivated by a concern over just how we might facilitate the "open-minded and mutually respectful dialogue" that Hanson (2008: 41) calls for between area specialists and those who are more partial to general theories or deductive models. Below, I examine the prospects of getting such a dialogue off the ground in view of the prevailing methodological currents and epistemological divides in the discipline. I suggest that rival epistemological orientations not only drive the debates over the status of area studies, as Hanson suggests, but also generate quite different types of area-focused scholarship some of which have a better chance than others of productively engaging general comparativists in the discipline. To facilitate even this more narrowly circumscribed dialogue, which I view as a pre-condition for a wider conversation encompassing more varied perspectives, a useful mediating role can be played by *cross-regional* small-N comparison (as opposed to area-bound comparative studies or single-country studies). I consider some reasons why this approach, in spite of its storied past, seems to have been relegated to the margins of the field given the character and dynamics of the methodological debates of the 1990s. I argue that cross-regional small-N comparison, despite its limitations, continues to offer a distinctive type of analytic leverage long associated with Millean comparative analysis. More importantly, it can help area specialists better defend their place in the field by judiciously engaging their contributions and making their empirical and theoretical value more obvious to comparativists focused on general theories and models.[1]

The Value and Heterogeneity of Area Studies: The Significance of Epistemological Divides

Most area specialists participate actively in interdisciplinary communities and invest much time and effort in gaining language skills and expertise to carry out fieldwork in particular countries or regions. These interactions and investments do not, however, automatically disqualify a scholar from being able to participate actively in general theoretical and methodological debates in comparative politics. This is not a zero-sum game. It is possible for area specialists to make theoretically significant contributions to the discipline while continuing to engage other area specialists affiliated with anthropology, sociology or more humanistic disciplines. This dual role makes sense in view of Hanson's (2008) observation that the geographic boundaries of areas often coincide with the analytic boundaries appropriate for many of the big questions in the field, such as those linked to the long-term institutional legacies of fallen empires or the spatial diffusion of particular societal attributes. To this, I would add that a number of analytically useful concepts in comparative politics owe much to the work of scholars who were trained as area specialists and whose research focused on particular countries or regions.[2] Thus, far from being relegated to the position of data providers (Bates 1996), area specialists should be recognized as contributing to the richness and expansiveness of comparative politics, simultaneously engaging (or critiquing) general theories in the field while helping to build bridges between comparative politics and other fields and disciplines.

It is worth emphasizing, however, that area-focused scholarship does not constitute a uniform approach. While the debate over area studies reflects "a struggle over the validity of rival theoretical paradigms" (Hanson 2008: 38), the epistemic commitments associated with these rival paradigms also contribute to the heterogeneity of research products that fall under the rubric of area studies. Those who invest in area expertise for the purpose of testing or crafting general hypotheses, for example, proceed from foundations that are quite different from those seeking a deeper understanding of particular context-specific discourses and practices. Recent discussions over methods highlight the value of process-tracing through careful analysis of particular cases (e.g. George and Bennett 2005; Gerring 2007; Goertz 2006). However, prevalent conceptions of standards and research design principles in case study approaches (Brady and Collier 2004), in spite of their openness to different methodologies, continue to reflect a broadly empiricist foundation not fundamentally different from that informing the conventional quantitative worldview (McKeown 2004). In fact, the very act of treating one's object of empirical analysis as a "case" and one's observations as "data points" suggests a commitment to a nomothetic endeavor ultimately geared towards identifying or confirming general laws or law-like regularities. There is nothing inherently problematic about such an endeavor. But, it is not the only way to think about area-focused scholarship. In fact, the prospects for fruitful dialogue between area specialists and generalists depend on the latter's awareness of the varied objectives and foundations associated with the work undertaken by the former.

As I have argued elsewhere (Sil 2000), the nomothetic-idiographic divide is not so much a dichotomy as a spectrum reflecting different degrees of sensitivity to context matched with different levels of ambition with respect to generalization. Thus, area studies encompass many distinct forms of nomothetic and idiographic scholarship—ranging from illustrations of theoretical models and case studies to historical narratives and ethnography—each reflecting distinctive epistemological principles. Area-specific narratives can certainly support nomothetic endeavors, and interpretive understandings can certainly be constructed as causal stories (Abbott 2004; Bevir 2006). But, area-focused scholarship that is self-con-sciously idiographic in its orientation is premised on the fundamental subjectivity of both social life and social inquiry and is thus less concerned with general patterns and regularities than with illuminating the processes of "meaning making" (Yanow 2006) among actors engaged in shared practices within their respective contexts. In fact, the epistemological chasm between two single-country studies, one designed as a case study to examine a preexisting hypothesis and one employing ethnography to better understand certain local practices, is at least as wide as that between a small-N comparison and a large-N study employing multiple regression analysis.

The heterogeneity of area studies has crucial implications for the scope and character of whatever dialogue might be possible between area specialists and general comparativists. Some kinds of area-based inquiry are better positioned (epistemologically speaking) to engage in productive conversations with generalists in comparative politics. Specifically, area specialists who accept the basic objectives of nomothetically oriented scholarship (for example, those who proceed from empiricism, realism or even pragmatism) are more likely than area specialists closer to the idiographic end (for example, those who embrace hermeneutics, phenomenology or post-structuralism) to be able to engage in fruitful dialogue with each other and with general comparativists concerned with universal theories or deductive models. Dialogue across a wider epistemological divide is certainly possible and desirable. I fear, however, that the prospects for such a dialogue cannot be very good when groups of scholars who are closer together in epistemological terms still manage to speak past one another on a regular basis.

One reason why even a more narrowly circumscribed conversation between area specialists and generalists has been difficult to sustain, I suggest, is the absence of a mediating third-party in the form of *cross-regional* small-N comparative studies. Once a prominent part of the repertoire of comparative politics, cross-regional comparisons appear to have become relegated to the margins. Below, I consider why this has happened and argue for the revival of cross-regional small-N comparison not only because of its previously recognized analytic benefits but also because of its potential role in facilitating a productive dialogue between area specialists and generalists in the discipline.

The Current Predicament of Cross-Regional Small-N Comparison

Mahoney and Rueschemeyer (2003) note that comparative historical analysis has regained a place at the center of the social sciences. The overwhelming majority of the examples they cite (Mahoney and Rueschemeyer 2003: 3–4, fn1–8), however, are either historical studies of single countries or comparisons of countries within particular areas (most frequently, Western Europe, Latin America and the Middle East). Of the fifty or so single-authored books they take note of, barely a half dozen involve comparisons of cases drawn from different areas (e.g., Evans 1995; Lustick 1993; Marx 1998; Seidman 1994; Waldner 1999). Evelyne Huber (2003: 1), President of APSA's comparative politics section at the time, laments: "Systematic cross-regional comparison is neither widely practiced nor written about from a methodological point of view, in contrast to case studies, small-N comparative historical studies within the same region or where region does not figure as a variable." Examining thirty of the most widely assigned books in comparative politics seminars and reading lists, Huber finds only four that feature systematic cross-regional comparisons based on empirical evidence.

All things being equal, the proportion of small-N comparative studies that feature cases drawn from different regions should be much higher than is presently the case given that the rationale for doing small-N comparative analysis has to do with the analytic leverage provided by carefully matched "most similar" and "most different" cases rather than with commonalities inherent within geographic boundaries (Przeworski and Teune 1970; Gerring 2007; Tarrow, forthcoming). It is true that those commonalities can provide crucial controls for "most similar" systems comparisons for certain kinds of questions, but not nearly as frequently as suggested by the ratio of area-bound to cross-regional small-N studies over the past ten to fifteen years. It is also worth noting that those who do produce small-N studies have tended to highlight their familiarity with specific regions rather than the appropriateness of the set of cases chosen to analyze a given problem. Thus, cross-regional small-N comparative analysis as well as the rationale historically invoked to justify its use are both underrepresented in the field at present.

I can only speculate as to why this is so. It is likely that the interplay of debates over quantitative and qualitative researchers (e.g. "The Quantitative-Qualititative Disputation" 1995; Brady and Collier 2004) and the debates over the theoretical standing of area studies (e.g. Bates, Johnson, and Lustick 1997) has resulted in cross-regional small-N comparison being buffeted from multiple directions. On the one hand, area specialists have responded to questions about their commitment to rigor and theory by emphasizing their main comparative advantage: the detailed analysis of sources and data about particular times and places with skills that they alone possess. This has meant challenging the merits of country-specific interpretations included in broader comparative studies produced by scholars without the credentials and training to do fieldwork in particular countries. In this regard, it has become fashionable to cite Theda Skocpol's (1979) book as an example of the difficulties that result from the unreflective use of secondary sources in her case studies (e.g. Burawoy 1989; see also Lustick 1996). This kind of criticism has been reinforced by challenges to the general value of "huge comparisons" focused on "big countries" (Tilly 1984) as opposed to more "close range" analysis based on fieldwork in particular locales (Collier 1999).

From the other side, those interested in hypotheses expected to be valid across regions have tended to rely on increasingly sophisticated techniques of statistical analysis or formal modeling. Critics of Millean small-N comparative analysis have noted the built-in limitations of efforts to construct generalizations on the basis of a few cases, emphasizing especially the problem of too few cases and too many variables (Goldthorpe 1997; Lieberson 1991). But, rather than consider these limitations alongside the distinctive payoffs of small-N comparison (and all methods involve tradeoffs reflecting both limitations and payoffs), proponents of case study research have tended to emphasize the value of process-tracing and causal narratives within single cases (George and Bennett 2005; Mahoney 2003) or occasionally the use of parallel case studies of a few similar countries or locales (Seawright and Gerring 2008). In practice, case studies understood in this way have been incorporated into multi-method approaches alongside formal or statistical models. This is true in the case of the various examples of "analytic narratives" assembled by Bates et al (1998); it is also true of many studies where statistical analyses have been juxtaposed with case studies employed to construct an alternative argument (e.g. Herrera 2005). In sum, cross-regional small-N comparison lags far behind statistical and formal approaches when it comes to single-method research, and has only rarely been employed in multi-method research.

I want to emphasize that it is specifically *cross-regional* comparisons that seem to have been put most on the defensive by the thrusts and parries of the opposing sides in recent methodological debates. For those who prefer statistical analysis or formal models, there may be a benefit to demonstrating one's familiarity with one or a few cases in the eyes of a particular area studies audience. For this purpose, little is gained by doing cross-regional comparative analysis. For area specialists seeking to demonstrate their breadth and engagement with the discipline, there are certainly benefits of adding even a single additional case when it comes to fending off the challenge that a given narrative is idiosyncratic (Rueschemeyer 2003). At the same time, there are fewer risks when the additional case is not viewed by the field as requiring much stretching of one's primary area of expertise. In both scenarios, those partial to small-N comparison have reacted to recent methodological debates more by demonstrating their in-depth familiarity with a given case or region than by insisting on the distinctive analytic leverage offered by "paired comparison" as Tarrow (forthcoming) has recently sought to do.

Significantly, for those who do venture beyond a single case but not a single area, some areas appear to be more suitable than others for deploying the kinds of expertise and skill-sets one associates with trained area specialists. Area-bound comparative analyses appear to be more common in areas where plentiful standardized data and translated archival material is available (as in Western Europe) or where a single language can support fieldwork in several countries (as in Spanish-speaking Latin America and the Middle East). In other areas, however, small-N comparisons involve neither a single set of language skills nor standardized or translated primary material given the linguistic or socio-cultural heterogeneity of the areas. This is true, for example, for countries within East Asia, Eastern Europe, and Africa (unless one focuses on narrower groups of countries with a shared lingua franca inherited from a common colonizer). In none of the latter regions would mastery of a single language help with conducting comparative analysis across several countries, thus necessitating either a second or third language or reliance on secondary sources and support provided by native scholars or country experts.[3] And, yet, a comparative study of China, Japan, Korea and Vietnam—or of Poland, Hungary, Romania and Bulgaria—is assumed to have the same seal of approval in terms of area expertise as a comparative study of a few Middle Eastern or Latin American states based on primary research using Arabic or Spanish. The point is this: certain kinds of small-N studies are presumed to be more tolerable in the present intellectual climate because they are area-bound even though their research requirements (particularly in relation to language skills and ongoing familiarity) are not fundamentally different from those of cross-regional small-N comparisons.

Of course, the fact that area-bound small-N studies do not always involve systematic use of languages and primary sources does imply that, in both analytic and practical terms, there *is* space in the field for cross-regional comparative analysis carried out without the expectation that the same level of language proficiency or country expertise will be leveraged in each case. If this is true, then there is no reason, other than seeking to consolidate one's credentials as an area specialist or to engage a specific area-studies audience, for limiting a small-N study to a single area unless the research question specifically requires this. Certainly, increased familiarity with one's cases is always helpful for a small-N study, and it is possible that area-bound small-N studies increase the degree of familiarity with cases even if language skills are not systematically employed to analyze comparable primary sources. It is also true that cross-regional comparisons probably magnify the challenges posed by the multiplicity of historical narratives and secondary sources pertaining to any chosen country (Lustick 1996). However, there is no inherent reason, other than the pressure to publish more quickly, why one cannot patiently increase one's familiarity with cases through careful attention to the nuances of each case, deep immersion in the relevant literature, and judicious handling of the relevant sources.[4]

The Mediating Role of Cross-Regional Comparison

I am by no means suggesting that small-N comparative analysis is inherently superior to other approaches, or that cross-regional comparison is inherently better than area-bound comparative analysis. Small-N comparative analysis does have built-in limitations, most notably the aforementioned problem of too few cases and too many variables (Lieberson 1991). However, all methods involve tradeoffs that require a balanced consideration of shortcomings alongside the potential payoffs (Sil 2000). The distinctive analytic payoffs of small-N comparison, which other approaches are not designed to yield, include: aggregating findings about relative strengths of particular causal mechanisms, recognizing non-linear complex interactions among multiple causal factors, identifying the origins and trajectories of pathways across comparable contexts, generating or probing novel hypotheses about clearly delimited problems, and adapting and fitting concepts in ways that help to negotiate the particularities of locales (George and Bennett 2005; Gerring 2007; Goertz 2006; Mahoney 2003; Rueschemeyer 2003; Seawright and Gerring 2008; Tarrow forthcoming).

Certainly, these benefits can be extracted from area-bound small-N comparison when the problem is area-specific as in the case of a study analyzing the impact of European Union membership on social policy or economic liberalization in Eastern Europe.[5] For many other questions that comparativists care about, however, bounding case selection within geographic areas is likely to weaken the kind of analytic leverage small-N comparison is designed to provide if for no other reason but that it artificially truncates the range of variation that could be represented by the full population of relevant cases (Geddes 2003: 97; Goertz and Mahoney 2006: 186). This is the case where the problem under investigation is present across areas as in the case of, say, the relationship between the robustness of democracy and the level or pace of economic development, the implications of high or low union density on social policy in developing countries, or the likelihood of violence in multiethnic nations employing various electoral systems. For such problems, large-N studies featuring multiple regression analysis can certainly be employed to provide a probabilistic assessment based on correlational estimates, but small-N comparisons can provide different kinds of benefits, particularly in relation to tracing complex causal interactions and identifying whether and how certain types of mechanisms contribute to convergent and divergent pathways of change (Rueschemeyer 2003). These benefits may be diluted when case selection is governed more by considerations of familiarity with a given area than by considerations of the relevant populations of cases (Goertz and Mahoney 2006).

Yet, there are a handful of recent books authored by scholars who have self-consciously pursued small-N comparison without confining themselves to geographically defined areas even if they have more familiarity with a given country or region. Lieberman's (2003) paired comparison of the politics of race and taxation in Brazil and South Africa, for example, is premised on the fact that his chosen pair of cases provide the most relevant controls for the puzzle even if they are not from the same region. Lieberman (2003: 34–35) notes: "Although the countries are situated in different world regions, different languages are spoken there, and some other obvious differences exist, the selection of South Africa and Brazil as the foundation for their analysis is due largely to their great similarities in social, economic, and geopolitical terms." In a similar vein, Kohli (2004) goes beyond India, the country with which he is most familiar, to compare how colonial legacies and related patterns of state construction in Brazil, India, Nigeria and South Korea generated distinctive patterns of state intervention and varying levels of success in promoting industrialization. Chen's (2007) comparative study of post-communist nationalism stands out in its effort to generate a comparative-historical study of four countries (Hungary, Romania, Russia and China) that are not confined to a single area but do enable her to capture the full range of variation in the extent of liberalism infusing nationalist discourse in formerly communist countries. And, Solinger (2009), a China-specialist by training, and Hanson (forthcoming), a Russia-specialist by training, both go beyond their "home turfs" to generate ambitious cross-regional studies that take advantage of a geographically diverse set of matched cases to explore their respective puzzles. Solinger compares France, Mexico and China to analyze the pressures faced by labor in different types of economies; Hanson compares post-Soviet Russia, Weimar Germany, and Third-Republic France to analyze the linkages between ideology and party-formation. In all of these cross-regional comparative studies, scholars have chosen to forego the benefits provided by their initial area expertise in order to better leverage the intended benefits of "most similar" and/or "most different" system designs (Przeworski and Teune 1970; George and Bennett 2005; Tarrow forthcoming).

Less commonly acknowledged but equally significant are the *dialogical* benefits that cross-regional small-N studies have historically brought to the table. One of the most valuable functions performed by the classic exemplars of cross-regional small-N comparison (e.g. Moore 1966; Skocpol 1979) was the role they played in stimulating conversations that cut across different communities of area specialists while also linking area-specific arguments to general arguments about big questions in comparative politics. While one can critique the reasoning behind the inferences drawn and the use of sources in the case studies in these classic works, one cannot deny that these studies have left a lasting impression on the field of comparative politics. By providing similarly structured interpretations of cases drawn from different regions, these works have served as a common pivot point for discussions among separate communities of country or area specialists. And, they have offered generalizations that could be treated as spurring further research among future generations of scholars, or at least providing foils for developing more nuanced hypotheses. It is this dual communicative function—operating "horizontally" across communities of area specialists and "vertically" between generalists and area studies writ large—that makes cross-regional small-N comparison useful in a way that a within-case analysis or an area-bound comparative study cannot be.

This is precisely why I believe that creating space once more for cross-regional comparative analysis is in the interest of area specialists and of the field of comparative politics writ large. From the perspective of area specialists, cross-regional comparison, rather than being the enemy as was often the case in the past, can serve as a valuable ally, demonstrating theoretically interesting parallels or differences in the debates and studies generated within separate communities of area scholars. In engaging the case-specific arguments advanced as part of a cross-regional project, different communities of area specialists can also join forces in defending the value of their work to skeptics in the discipline. From the perspective of general comparativists, small-N comparisons provide a useful stepping stone in identifying crucial mechanisms or illuminating important contextual variables before plunging into efforts to apply general models or test hypotheses through studies focused on single countries. In short, cross-regional comparative approaches serve as a valuable point of engagement and translation between general concepts and theories in the field of comparative politics and the more specific (but often comparable) debates among various communities of country- or area- specialists.

Conclusion

Area-focused scholarship of all varieties—ranging from case studies intended to explore general hypotheses to context-bound narratives generated through ethnography or discourse analysis—have long contributed to the richness and vibrancy of the field of comparative politics, sometimes generating enduring concepts, sometimes reinforcing or challenging prevailing general theories. The question for comparativists is not whether area studies ought to have a place in the field of comparative politics but rather, as Hanson (2008) suggests, how to generate productive conversations between area studies and scholars working on general theories and models. Such conversations are not going to be jump-started by simply noting the value of area-focused scholarship. This is precisely why I have focused on the mediating role of cross-regional comparative research in highlighting the connections between different clusters of area-focused scholarship and general theoretical and methodological debates in the field.

Unfortunately, the dynamics of methodological debates are such that those partial to small-N comparison have typically preferred to go closer to the ground, seeking the cover of area-specific expertise rather than reaffirming the analytic leverage small-N analysis is designed to provide. In light of this state of affairs, it is worth applauding the handful of studies that have boldly embarked on cross-regional comparison, recognizing the limitations of this approach but standing fast in highlighting its distinctive payoffs. But, the broader argument for reviving cross-regional small-N comparison has to do with its distinctive dialogical benefits—that is, the role it can play in bridging the gulf that continues to separate area-studies communities from generalists in comparative politics. If this gulf is not reduced, area specialists could find themselves further misunderstood and marginalized, and the field of comparative politics will be impoverished as a result.

Notes

[1] This essay is part of a longer article (in progress) analyzing the past contributions and current utility of cross-regional small-N comparative analysis. I am grateful to Gary Goertz, Allison Evans and Eileen Doherty-Sil for comments and suggestions.

[2] The concept of "corporatism," for example, has long held an important place in comparative politics (Collier 1995; Wiarda 1997). Initially deployed in a social scientific context in Philippe Schmitter's (1971) single-country study of Brazil, it came to be employed in studies of postwar Europe (Katzenstein 1984; Schmitter and Lehmbruch 1979), post-socialist transitions (Connor 1996; Iankova 1998) and even the European Union (Gorges 1996). Similarly, the concept

of "consociationalism," initially constructed on the basis of a detailed study of the Netherlands (Lijphart 1968), proceeded to influence debates over the design of democratic institutions in divided societies a quarter century later (Gabel 1998; McGarry and O'Leary 1993). And, students of Eastern Europe have turned what was seen as a major failure of an area-studies community (the inability to predict the velvet revolutions of 1989) into an asset, using the varied origins and pathways of post-communist transformations to refine and flesh out such abstract concepts as historical legacies, path dependence, spatial diffusion, and temporal context (e.g. the essays in Ekiert and Hanson 2003).

[3] Very rarely, a scholar may have learned and used multiple languages for doing fieldwork in different countries within East Asia (e.g. Hoston 1997) or Eastern Europe (e.g. Orenstein 2001). But, this is more the exception than the rule among East Asianists and East Europeanists conducting comparative analysis across several countries.

[4] In my own paired comparison of Japan and Russia (Sil 2002), I supplied an appendix that reviewed the main lines of debate in Japanese and Russian studies on the substantive issues I was contending with. The point was not only to demonstrate that I had sampled a wide range of perspectives in my handling of secondary sources, but also to demonstrate how the two case studies were similarly positioned in each case vis-à-vis the existing historiographic traditions (see Appendix B). This is one way to reduce the dangers of selection bias in dealing with multiple narratives in comparative studies (Lustick 1996).

[5] However, as noted elsewhere (Chen and Sil 2007), there are many broader questions related to post-communist transformations where it is worth designing comparisons that extend beyond Eastern Europe to China, Vietnam, and even sub-national units such as the state of West Bengal in India.

References

Abbott, Andrew. 2004. *Methods of Discovery*. New York: W. W. Norton.

Bates, Robert. 1996. "Letter from the President: Area Studies and the Discipline." *APSA-CP: Newsletter of the APSA Comparative Politics Section* 7:1 (Winter), 1–2.

Bates, Robert, Chalmers Johnson, and Ian Lustick. 1997. Contributions to "Controversy in the Discipline: Area Studies and Comparative Politics." *PS: Political Science* 30:2, 166–179

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

Bevir, Mark. 2006. "How Narratives Explain." In *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*. Dvora Yanow and Peregrine Schwartz-Shea, eds. (Armonk, NY: M. E. Sharpe), 281–290.

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

Burawoy, Michael. 1989. "Two Methods in Search of Science: Skocpol versus Trotsky." *Theory and Society* 18, 759–805

Chen, Cheng. 2007. *The Prospects for Liberal Nationalism in Post-Communist States*. University Park: Pennsylvania State University Press.

Chen, Cheng and Rudra Sil. 2007. "Stretching Postcommunism: Diversity, Context, and Comparative Historical Analysis." *Post-Soviet Affairs* 23:4, 275–301.

Collier, David. 1995. "Trajectory of a Concept: 'Corporatism' in the Study of Latin American Politics." In *Latin America in Comparative Perspective*. Peter Smith, ed. (Boulder: Westview Press), 135–162.

Collier, David. 1999. "Data, Field Work and Extracting New Ideas at Close Range (Letter from the President)." *APSA-CP: Newsletter of the APSA Organized Section in Comparative Politics* (Winter), 1–4.

Connor, Walter D. 1996. *Tattered Banners: Labor, Conflict and Corporatism in Post-Communist Russia*. Boulder: Westview.

Ekiert, Grzegorz and Stephen E. Hanson eds. 2003. *Capitalism and Democracy in Central and Eastern Europe*. New York: Cambridge University Press.

Evans, Peter. 1995. *Embedded Autonomy: States and Industrial Transformation*. Princeton: Princeton University Press.

Gabel, Matthew J. 1998. "The Endurance of Supranational Governance: A Consociational Interpretation of the European Union." *Comparative Politics* 30:4, 463–475.

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

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

Gerring, John. 2007. *Case Study Research: Principles and Practices*. New York: Cambridge University Press.

Goertz, Gary. 2006. *Social Science Concepts: A User's Guide*. Princeton: Princeton University Press.

Goertz, Gary and James Mahoney. 2006. "Negative Case Selection: The Possibility Principle." In Gary Goertz, *Social Science Concepts: A User's Guide*. (Princeton: Princeton University Press), 177–210.

Goldthorpe, John. 1997. "Current Issues in Comparative Methodology." *Comparative Social Research* 16, 1–2.

Gorges, Michael. 1996. *Euro-Corporatism? Interest Intermediation in the European Community*. Lanham, MD: University Press of America.

Hanson, Stephen E. 2008. "The Contribution of Area Studies." *Newsletter of the APSA Section on Qualitative & Multi-Method Research* 6:2, 35–43.

Hanson, Stephen E. Forthcoming. *Ideology, Uncertainty, and Democracy: Party Formation in Third Republic France, Weimar Germany, and Post-Soviet Russia*. Cambridge: Cambridge University Press.

Herrera, Yoshiko. 2005. *Imagined Economies: The Sources of Russian Regionalism*. New York: Cambridge University Press.

Hoston, Germaine. 1997. *State, Identity and the National Question in China and Japan*. Princeton: Princeton University Press.

Huber, Evelyn. 2003. "The Role of Cross-Regional Comparison (Letter from the President)." *APSA-CP: Newsletter of the APSA Organized Section in Comparative Politics* 14:2, 1–6.

Iankova, Elena. 1998. "The Transformative Corporatism of Eastern Europe." *Eastern European Politics and Societies* 12:2, 222–264.

Katzenstein, Peter. 1984. *Corporatism and Change: Austria, Switzerland and the Politics of Industry*. Ithaca: Cornell University Press.

Kohli, Atul. 2004. *State-Directed Development: Political Power and Industrialization in the Global Periphery*. New York: Cambridge University Press.

Lieberman, Evan S. 2003. *Race and Regionalism in the Politics of Taxation in Brazil and South Africa*. New York: Cambridge University Press.

Lieberson, Stanley. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70:2, 307–320.

Lijphart, Arend. 1968. *The Politics of Accommodation: Pluralism and Democracy in the Netherlands*. Berkeley: University of California Press.

Lustick, Ian. 1993. *Unsettled States, Disputed Lands: Britain and Ireland, France and Algeria, Israel and the West Bank-Gaza*. Ithaca: Cornell University Press.

Lustick, Ian. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias." *American Political Science Review* 90:3, 605–618.

Mahoney, James. 2003. "Strategies of Causal Assessment in Comparative Historical Analysis." In *Comparative Historical Analysis in the Social Sciences*. James Mahoney and Dietrich Rueschemeyer, eds. (New York: Cambridge University Press), 337–372.

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

Marx, Anthony. 1998. *Making Race and Nation: A Comparison of South Africa, the United States, and Brazil*. Cambridge: Cambridge University Press.

McGarry, John and Brendan O'Leary, eds. 1993. *The Politics of Ethnic Conflict Regulation: Case Studies of Protracted Ethnic Conflicts*. London: Routledge.

McKeown, Timothy. 2004. "Case Studies and the Limits of the Quantitative Worldview." In Henry Brady and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. (Lanham, MD: Rowman & Littlefield), 139–167.

Moore, Barrington. 1966. *Social Origins of Dictatorship and Democracy*. Boston: Beacon.

Orenstein, Mitchell. 2001. *Out of the Red: Building Capitalism and Democracy in Post-Communist Europe*. Ann Arbor: University of Michigan Press.

Przeworski, Adam and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience.

Rueschemeyer, Dietrich. 2003. "Can One or a Few Cases Yield Theoretical Gains?" In *Comparative Historical Analysis in the Social Sciences*. James Mahoney and Dietrich Rueschemeyer, eds. (New York: Cambridge University Press), 305–336.

Schmitter, Philippe C. 1971. *Interest Conflict and Political Change in Brazil*. Stanford: Stanford University Press.

Schmitter, Philippe and Gerhard Lehmbruch, eds. 1979. *Trends Towards Corporatist Intermediation*. Beverly Hills: Sage Publishers.

Seawright, Jason and John Gerring. 2008. "Case-Selection Techniques in Case Study Research." *Political Research Quarterly* 61, 294–308.

Seidman, Gay. 1994. *Manufacturing Militance: Workers' Movements in Brazil and South Africa, 1970-1985*. Berkeley: University of California Press.

Skocpol, Theda. 1979. *States and Social Revolutions*. New York: Cambridge University Press.

Sil, Rudra. 2000. "The Division of Labor in Social Science Research: Unified Methodology or 'Organic Solidarity'?" *Polity* 32:4, 499–531.

Sil, Rudra. 2002. *Managing 'Modernity': Work, Community and Authority in Late-Industrializing Japan and Russia*. Ann Arbor: University of Michigan Press.

Solinger, Dorothy. 2009. *States' Gains, Labor's Losses: China, France, and Mexico Choose Global Liaisons, 1980-2000*. Ithaca: Cornell University Press.

Tarrow, Sidney. Forthcoming. "The Strategy of Paired Comparison: Towards a Theory of Practice." *Comparative Political Studies*.

Tilly, Charles. 1984. *Big Structures, Large Processes, and Huge Comparisons*. New York: Russell Sage.

Waldner, David. 1999. *State Building and Late Development. Ithaca: Cornell University Press*. Ithaca: Cornell University Press.

Wiarda, Howard. 1997. *Corporatism and Comparative Politics: The Other Great 'Ism.'* Armonk, NY: M. E. Sharpe.

Yanow, Dvora. 2006. "Thinking Interpretively: Philosophical Presuppositions and the Human Sciences." In *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*. Dvora Yanow and Peregrine Schwartz-Shea, eds. (Armonk, NY: M. E. Sharpe), 5–26.

Symposium: Ethnographic Methods in Political Science

Introduction

Edward Schatz
University of Toronto, Canada
ed.schatz@utoronto.ca

Ethnography has long been in the arsenal of approaches available to the political scientist, but recent discussions of qualitative methods sometimes equate ethnography with narrative, with case-studies, with interviews, or with fieldwork (broadly understood). There may be family resemblances and elective affinities among some of these approaches and techniques, but ethnography has its own distinctive intellectual history and its own potential value for the study of politics.

This symposium started to take form in 2006, with a workshop in Toronto that channeled many conversations about ethnography that were already ongoing in the discipline and ultimately resulted in a recent volume (Schatz 2009). In 2009 we sought to push our conversations about ethnography and political science further during a roundtable at APSA (returning, full-circle, to Toronto). Buoyed by these efforts, as well as other recent work (Joseph, Mahler, and Auyero 2007, Wedeen in press), we use the space below to chart out some of the distinctive features of ethnographic approaches to politics, in the hope that doing so will enable us better to understand the value that they bring to political research. We pepper our discussion with illustrations of ethnography's "added value"—for studies of postcommunism, for studies of civil society, for understanding political contingency, for understanding lived history, and for understanding the discipline of political science itself.

Second, we ask a series of questions—about what constitutes a research "site," about a researcher's positionality, about forms of intellectual discovery, about "post-hocism," about problem-driven research, and about macro-developments in the discipline—that an ethnographic sensibility brings to the fore. These are questions about which QMMR members are, no doubt, aware; ethnography simply puts them into the starkest relief.

The larger purpose of the symposium is to ask what a *political* ethnography should look like. That is, how should ethnographic methods and approaches be adapted best to meet the needs of political research? We provide no definitive answers, but this is a conversation that, we deeply believe, needs to occur.

References

Joseph, Lauren, Matthew Mahler, and Javier Auyero, eds. 2007. *New Perspectives in Political Ethnography*. New York: Springer.

Schatz, Edward, ed. 2009. *Political Ethnography: What Immersion Contributes to the Study of Power*. Chicago: University of Chicago Press.

Wedeen, Lisa. In Press. "Ethnography and Political Science." *Annual Review of Political Science* 13.

What's Political About Political Ethnography? Abducting Our Way Toward Reason and Meaning

Dvora Yanow
Vrije Universiteit, Amsterdam
d.yanow@fsw.vu.nl

"This is an unusual political science dissertation; it would be more at home in an anthropology or sociology department."

— Internal Examiner, a political scientist who does survey research, at a recent dissertation defense, referring to an ethnographic study of campaign organizers

As the discipline re-engages with its earlier history of ethnographic approaches to the study of political matters (e.g., Selznick 1949, Blau 1955, Kaufman 1960, Pressman and Wildavsky 1973, Lipsky 1980), we need to engage two dialogues.[1] One is a conversation with those parts of the discipline not familiar with it concerning what ethnography is—a more general conversation that draws on works outside of political science, as well. The other is a narrower, or perhaps deeper, conversation among "ourselves," to elucidate the characteristics of *political* ethnography and to map its terrain, especially as political ethnography can be done from both objectivist-realist and constructivist-interpretivist ontological and epistemological positions (Yanow and Schwartz-Shea 2006). Parallel conversations are taking place in some subfields, such as public policy concerning what policy ethnography is, and in other disciplines, such as organizational studies concerning the character and practice of organizational ethnography. Here, I want to engage two of the issues that, for me, lie at the center of what we need to be exploring in political and policy ethnography today: its distinctive characteristics; and its underlying logic of inquiry. Getting to those points requires some definitional brush-clearing with respect to what this "ethnography" thing is, and in the process I collapse these two conversations (as well as whatever distinctions might exist between political and policy ethnography: I will largely subsume the second under the first).

What is Political/Policy Ethnography?

It is not only that political ethnography brings something useful to the study of the political; it does so in ways that are different from "anthropological ethnography."

"Ah," you say, "but ethnography comes from anthropology; how can it be different?"

It is time to let go of this notion of ethnography's origins. As the history of anthropology narrated by social anthropologist Oscar Salemink (2003) claims, ethnographic methods did not originate with the academic discipline of anthropology. Their origins lie in *administrative* practices; that is, in empires' needs to manage far-flung and distant outposts—colonial ones, to be sure, with all the paternalism and 'Orientalism' (Saïd, 1978) and racism (and sexism and able-bodiedism, still largely unspoken of) they entailed, which marked those methods in ways their users are still contending with; but administrative nonetheless, an *organizational* and *political* practice after all! Salemink makes the point this way:

> contrary to the now common assumption that ethnography is the descriptive (or even the field research) part of anthropology...professional anthropology is a fairly recent manifestation of ethnographic practice.... Missionaries, military explorers, colonial administrators, plantation owners, development workers, counterinsurgency experts, government officials, politicians, indigenous leaders—male and female—construct[ed] ethnographic images of these indigenous groups...according to their experiences and in order to suit their interests. Those ethnographic representations interacted with the ones of professional anthropologists, who created their representations in dialogue with, sometimes in opposition to, but always against the backdrop of those 'non-professional' ethnographic representations.[2] (2003: 2–3)

This is not to say that missionaries, development workers, and so on practiced ethnography in identical ways—but then, professional anthropologists today do not all practice in the same ways either. Nor does this deny that some current practices are much more attuned to the political dimensions of their practices and reflective about their roles in the construction of knowledge than, say, government officials were at the time. Even within academic anthropology, today's sensitivities toward emic/etic roles and knowledge, relational methods, reflexivity, and positionality have moved beyond what Malinowski engaged in calling for *participant* observation as the basis of its practice (cf. Salemink 2003, on the "myth" of Malinowski's invention of modern anthropological ethnography). It is, says Salemink, "better to regard academic anthropology as a specific instance of ethnographic practice than the other way around" (Salemink 2003: 10).

That enables those of us practicing political ethnography to stop looking over our shoulders for some imagined anthropological "ethnography police" who are going to arrest us for applying an inferior version, for leaving something out or getting something wrong. We need to let go of those anxieties. That will free us to work on delineating what marks the distinctiveness of *political* ethnography. Answering this draws on what it shares with other forms of ethnography, such as is practiced in organizational, educational or urban studies, as well as in anthropology. Methodologists increasingly distinguish three parts of a definition of ethnography, whether it is done in realist-objectivist or in constructivist-interpretivist methodological fashion. First, it entails a set of methods that rest on "being there," interacting with people in some form and to some extent, living—in approximation—the lives they lead and participating in what they are doing. This has entailed, traditionally, talking and doing—participating, in a situational role—while at the same time observing in a researcher role (on the dual role of the participant-observer, see Gans 1976). Second, "ethnography" refers to a particular kind of research writing: one that is narrative in form, retaining the word-based character of the data the researcher has generated. One matter requiring ethnographers' attention is the extent to which we allow our writing to "neaten up the mess" that characterizes political (and other forms of) life (see Law 2004); another concerns who gets to hold the pen—i.e., to what extent are we prepared to treat situational members as equals (see Down and Hughes 2009). Third is what is increasingly being called an ethnographic "sensibility" (Pader 2006; Ybema et al. 2009). Somewhat difficult to explicate, this refers to a hermeneutic-phenomenological orientation toward social life: seeing it as being meaning-filled, where these meanings are embodied in and communicated through what might be called the 'underlife' (or back stage; Goffman 1959) of human acts, language (including written, oral, and nonverbal), and objects—their situationally common-sensical, unwritten, unspoken, everyday, tacitly known textures. Ethnography aims to makes these explicit through the use of a particular set of "tools" in the field and a particular form of engagement on the printed page: i.e., that sensibility infuses the three phases of research, field-, desk-, and text-work (Yanow 2000).

This tripartite definition—"tools" plus writing plus sensibility—presumes ethnography's emplaced-ness, something "being there" seeks conceptually to encapsulate. One hallmark of political ethnography is its "multi-sited" character. Traditionally, anthropological ethnography treated of one tribe, in one physical location. Because that location was typically remote, it was physically bounded in ways that sites of political ethnographic research are often not. That "multi-sited" is a marked term—the unmarked norm being simply "the field"—makes the point. As a policy ethnographer, for example, I follow the policy (or the decision/s leading to and/or implementing the policy)[3]; and this leads me to all sorts of locations, from legislative chambers to legislators' offices, from the latter to imple-mentors' offices, from there to the on-site locations where street-level bureaucrats meet "clients," and beyond. These spaces are not bounded by local or national borders; they are bounded by the map of actors, stakeholders, and onlookers interested in the policy's outcomes. The policy itself is my "field," much as an organization might be, or a professional practice. It is the research topic (or focus, or question), in other words, that bounds the research field, not a geographic border. Anthropologists have discussed the fetish of "the field" that characterizes many of their discussions; growing availability limitations on their traditional sorts of research sites, as well as the effects of globalization, have led them to turn increasingly to policy, organizational, and other forms of anthropology (Shore and Wright 1997, Wright 1994) that are "multi"-sited in similar ways, although the marked term remains.

A second feature that the tripartite definition brings out is the role played by documents, something an anthropologist studying a tribe in the bush would not expect to have to include. As a policy ethnographer, I look for all manner of documents related to the policy issue I am studying. If I am investigating legislative processes, then I'm looking for (sub-)committee reports, position papers, correspondence, memos, newspaper coverage, and the like. If I am studying agenda setting, I might also look at interest group webpages, op-ed essays, leaflets, etc. If I am analyzing policy implementation, I might be after the range of materials that organizations produce. In his study of Soviet and Russian political life, Hopf (2002) looked at novels, magazines, and other sort of "low data" (Weldes 2006). Mitchell (1991, 2002), Scott (1998), and Wood (2009) similarly draw on a wide range of types of material data, including objects.

The tripartite definition also enables us to say something about what ethnography—even political ethnography—is not: "ethnography" is not synonymous with interviewing. Ethnographers have long talked to people, although rarely designating this talk with the term "interview." Political ethnography, however, is also conducted with elites; and one can hardly plan on encountering the Prime Minister or the Secretary of Defense at the water cooler or in the market place. And so, unlike bush ethnographers—by which I mean those doing traditional anthropological ethnography, without any deprecation—we tend to set up appointments with such folk for conversations, whose formality we designate with the term "interview." But no amount of observing the outer office while waiting for that interview will turn such a study into an ethnography. Be a fly on the wall; shadow the Prime Minister; hang out long enough to become operationally conversant with the organizational culture of the place, and then you begin to have claims to doing an ethnography of, e.g., the practice of Prime Ministry. But this is much more than a one-off interview, or even that plus a follow-up or a set of single interviews with people in the office. It entails that "being there" and that sensibility which gain one conceptual access to the unwritten, unspoken, common sense, everyday, tacit knowledge of the "Prime Ministry operating manual." (Note: Here is where methodological differences between realist-objectivist and constructivist-interpretivist ethnography emerge. They concern the reality status accorded to the researcher's observations and whether that "being there" is treated as generating objective truths about the situation or [co-]constructed knowledge. Both sensibility and content of research writing are, therefore, also likely to be different. Space limitations preclude me from engaging these points further.)

Abductive Ways of Knowing: A Logic of Inquiry

Most of us are familiar with deductive and inductive reasoning: the former, generating testable concepts from theories (deducing the particular from the universal); the latter, developing general laws from observations of particular instances (inducing the universal from the particular). Methodologists are increasingly suggesting that ethnography (and other interpretive research) draws on another logic of inquiry: abductive reasoning (Agar in press, Locke et al. 2008, Van Maanen et al. 2007; see also Glynos and Haworth 2007 on retroduction, used synonymously). Articulated first, and at length, by US pragmatist Charles Saunders Peirce,[4] abductive logic begins by registering the presence of a puzzle, of something surprising, and then seeking to explicate it. In an ethnographic study, the puzzle or surprise can commonly derive from a tension between the expectations one brings to the field (based on one's "a priori knowledge," drawing on a classic Kantian and then phenomenological point about new understanding emerging from prior knowledge or experience) and what one observes and/or experiences. This is one of the reasons that "strangerness" is so important in generating ethnographic knowledge: being a stranger—holding on to that quality for as long as possible—is necessary in order to see as explicitly as possible what is taken-for-granted, common sense, tacitly known for situational members; yet approximating ever more closely that membership position is important for generating insider understanding of what is only puzzling to an outsider. One asks oneself, in other words, what conditions render an event, a word, a relationship more 'natural'—less surprising, less puzzling.

There are a number of implications of this way of reasoning—of knowing—that are highly characteristic of ethnographic research. For one, ethnographers try to retain an openness to the possibility of surprises, as well as to avoid the "rush to diagnosis" that prematurely closes down analytic possibilities. As things, acts, words, concepts, etc. that surprise do not arrive pre-labeled as such, attention must be paid. Note that marking something as surprising requires attending to the expectations and other prior knowledge one brings to the field. Second, we create new concepts, relationships, explanations to give an account of these surprises or puzzles. Abductive reasoning thereby enables us to speak more clearly about the relationship between theory and data and the researcher's dual role: the surprise comes from encounters in the field; the explanations may come from encounters in the literature. For political ethnography, neither is sufficient: there is a recursive and reiterative process between theoretical and field encounters. It is recursive in that we perform abduction within abduction within abduction, as one "discovery" leads to another. It is iterative in that the same logic is repeated over and over again. As Agar (in press, 8) says, "surprises never stop; just the time and money do." Third, this attention to expectations and giving accounts provides rationale for reflexivity, including on one's own positionality, both discussed further below.

On the methodological front, in departing from that which puzzles, abductive reasoning is a far cry from beginning with formal hypotheses. This explains something about ethnographic (and other modes of interpretive) research design: why it is so difficult to write research proposals that stipulate ahead of time, before immersion in the field, what we expect to find or the concepts we intend to explore ("measure," "test") in that research. Puzzles can only emerge from the field! And it helps explain why research designs must be flexible in order to respond to field conditions.

So, as Agar (in press) notes, the key question we should be asking and answering in crafting research designs (and advising those undertaking them) should be not "What is your research plan?" but "Where are you going to start looking?" and "Now where are you going to look?" I am reminded of the old joke about the drunk under the street lamp looking for his lost keys: they fell over *there*, but he is looking *here* because that's where the light is. Ethnographers do both, in an ever-widening set of concentric circles: in an oblique sort of looking, we might well start where the light—or access, or funding—is, even though we are interested in something outside this limelight; and we move our way, gradually, over "there," in the process pushing out the boundaries of light. This is a version of the hermeneutic circle (for further discussion, see Schwartz-Shea and Yanow, forthcoming).

A different analogy captures a different facet of ethnographic abduction. When you skip a stone over a pond's surface, it leaves ripples; the stone may then be nowhere to be found, but we still track those ripples (at least for a while longer), which are the impact or the impression the stone has had on the water. Ethnography tracks the residual 'behaviors' that we see when events—Stone Surprises—have impacts on people, places, things, acts: those things we wish to understand. The "stone-surprise" may no longer be visible, but we can surmise that a stone had been there when we see the ripples, and we can look to clarify aspects of the conditions it created as it passed. The "looking" is the talking, observing (and perhaps doing), and reading that comprise political ethnography; the clarification is the search to explain the puzzle of the surprising ripples.

Abductive reasoning itself does not require that one search for meaning, or that that meaning be context-specific, as Agar (in press, 9) notes. But ethnography (along with other methods used interpretively) does! It requires both iterative-recursive abduction *and* a focus on contextual meaning: this lamp post, not all lamp posts, not even *that* one over there. Ernie Zirakzadeh (2009) discovered this in the field: when, by happenstance, he ducked for cover from police shooting at Basque separatists and found himself sharing the doorway with activists from one ETA group, he was instantly taken for a sympathizer and found himself, that evening, the center of attention, whereas he had hitherto been scorned; but now, he became persona non grata among other ETA groups whose doorway-ducking he had not aped, from whose perspective he was no longer 'neutral.' A particular doorway, a particular self-protective impulse, made a world of difference.

Conclusion: Reflexivity, Positionality, and Personality

At the end of her wonderful methodological reflection, Carol Cohn (2006: 106–107) comments on the relationship between a researcher's own personal proclivities and her choice of methods. About herself, Cohn says that she is genuinely interested in others, temperamentally, a listener, conflict-avoidant, attentive to feelings, and compelled to honest openness about her views. Such reflexivity about one's own "positionality"—ranging from the characteristics of one's location in the field setting (see Pachirat 2009, Zirakzadeh 2009) to one's personal attributes (e.g., Shehata 2006, 2009) —in the service of transparency with respect to knowledge generation and knowledge claims is becoming increasingly de rigueur in interpretive methods across the disciplines.

But Cohn's comments point toward another aspect of our current methodological debates in political science. These have swirled over the problematic of methods-driven research, and a common stance has been that the choice of research question should determine the methods, rather than vice versa. The difficulty with this reply lies in its suggestion that such choices are made in an entirely objective, rational fashion, as if the researcher engages at his desk in "eeny-meeny-miny-moe" over a range of possibilities for framing a research question. What is omitted from this conceptualization is the notion that, perhaps, our proclivities for articulating research questions in one way or another, as well as our methods "orientations," are shaped by something akin to personality traits. This appears to be the case in other professions (e.g., that some medical students choose to practice anesthesiology because it does not require them to interact with patients). Why should the practice of political science be immune to the same sorts of influence?

Those of us who follow an ethnographic star may do so because we are more comfortable listening, perhaps even because we are more comfortable on the margins, observing—whether in a highly interactive research setting or in our disciplinary subfields (see Ed Schatz's contribution in this symposium). It seems to me that we *all* need to take this up, political ethnographers included. My hope is that further attention throughout the discipline to this and other issues in political ethnography will dispense with the kind of comment with which this essay opens. That is what is at stake in these conversations.

Notes

[1] Parts of this essay, in an earlier version, were presented as the keynote address to the New School for Social Research, Department of Politics' Conference, "Interpreting Politics, History and Society in the 21st Century" (1 May 2009). My thanks to Tim Pachirat and his colleagues and students there for engaging the ideas, to Ed Schatz for giving me the opportunity to pursue them further at the roundtable and here, and to Ed and my other roundtable colleagues for challenging me in ways that helped me be as clear as possible about what I am trying to say.

[2] Salemink later adds: "To a large extent, the professionalization of fieldwork in British anthropology depended on the tactical denigration of both missionary and administrative ethnographies" (2003: 9). He notes there Talal Asad's argument (in his 1973 *Anthropology and the Colonial Encounter*) that the crucial issue was "not the complicity of anthropologists with colonialism, but the location of anthropology in the colonial context."

[3] This echoes Latour's notion, in science studies, of following the "fact" (Latour and Woolgar 1988); see also Brandwein (2006).

[4] Peirce's ideas about abduction apparently changed between his early writings and his later ones (Benjamin Herborth, personal communication, Potsdam, 12 September 2009). What I present here characterizes the latter. Patrick Jackson (personal communication, Toronto, September 1, 2009) notes that we should be cautious in invoking Peirce's ideas for interpretive methodological purposes, given their origins in positivist thought. I suspect that Herborth's point explains Jackson's.

References

Agar, Michael. In press. "On the Ethnographic Part of the Mix: A Multi-Genre Tale of the Field." *Organizational Research Methods*.

Blau, Peter M. 1955. *The Dynamics of Bureaucracy*. Chicago: University of Chicago Press.

Brandwein, Pamela. 2006. "Studying the Careers of Knowledge Claims: Applying Science Studies to Legal Studies." In *Interpretation and Method: Empirical Research Methods and the Interpretive Turn.* Dvora Yanow and Peregrine Schwartz-Shea, eds. (Armonk, NY: M. E. Sharpe), 228–243.

Cohn, Carol. 2006. "Motives and Methods: Using Multi-Sited Ethnography to Study US National Security Discourses." In *Feminist Methodologies for International Relations*. Brooke Ackerly, Maria Stern, and Jacquie True, eds. (New York: Cambridge University Press), 91–107.

Down, Simon and Michael Hughes. 2009. "When the 'Subject' and the 'Researcher' Speak Together: Co-Producing Organizational Ethnography." In *Organizational Ethnography: Studying the Complexities of Everyday Life*. Dvora Yanow, et al., eds. (London: Sage), 83–98.

Gans, Herbert. 1976. "Personal Journal: B. On the Methods Used in this Study." In *The Research Experience*. M. Patricia Golden, ed. (Itasca, IL: F.E. Peacock), 49–59.

Glynos, Jason and David Howarth. 2007. *Logics of Critical Explanation in Social and Political Theory*. New York: Routledge.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor Books.

Hopf, Ted. 2002. *Social Construction of International Politics: Identities and Foreign Policies, Moscow, 1955 and 1999*. Ithaca: Cornell University Press.

Kaufman, Herbert. 1960. *The Forest Ranger: A Study in Administrative Behavior.* Baltimore: Johns Hopkins Press.

Latour, Bruno and Steve Woolgar. 1988. *Laboratory Life: The Construction of Scientific Facts*, 2nd ed. Princeton: Princeton University Press.

Law, John. 2004. *After Method: Mess in Social Science Research*. New York: Routledge.

Lipsky, Michael. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.

Locke, Karen, Karen Golden-Biddle, and Martha S. Feldman. 2008. "Making Doubt Generative: Rethinking the Role of Doubt in the Research Process." *Organization Science* 19:6, 907–918.

Mitchell, Timothy. 1991. *Colonising Egypt*. Berkeley: University of California Press.

Mitchell, Timothy. 2002. *Rule of experts*. Berkeley: University of California Press.

Pachirat, Timothy. 2009. "The Political in Political Ethnography: Reflections from an Industrialized Slaughterhouse on Perspective, Power, and Sight." In *Political Ethnography: What Immersion Contributes to the Study of Politics*. Edward Schatz, ed. (Chicago: University of Chicago Press), 143–162.

Pader, Ellen. 2006. "Seeing with an Ethnographic Sensibility: Explorations Beneath the Surface of Public Policies." In *Interpretation and Method: Empirical Research Methods and the Interpretive Turn.* Dvora Yanow and Peregrine Schwartz-Shea, eds. (Armonk, NY: M. E. Sharpe), 161–175.

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

Saïd, Edward. 1978. *Orientalism*. New York: Pantheon Books.

Salemink, Oscar. 2003. "Ethnography, Anthropology and Colonial Discourse." In *The Ethnography of Vietnam's Central Highlanders: A Historical Contextualization, 1850-1990*. (London: Routledge Curzon), 1–39.
Schwartz-Shea, Peregrine and Dvora Yanow. Forthcoming. *Interpretive Approaches to Research Design.* NY: Routledge.
Scott, James C. 1998. *Seeing Like a State*. New Haven: Yale University Press.
Selznick, Philip. 1949. *TVA and the Grass Roots*. Berkeley: University of California Press.
Shehata, Samer. 2006. "Ethnography, Identity, and the Production of Knowledge." In *Interpretation and Method: Empirical Research Methods and the Interpretive Turn.* Dvora Yanow and Peregrine Schwartz-Shea, eds. (Armonk, NY: M. E. Sharpe), 244–263.
Shehata, Samer. 2009. *Shop Floor Culture and Politics in Egypt*. Albany: SUNY Press.
Shore, Cris and Susan Wright, eds. 1997. *Anthropology of Policy: Critical Perspectives on Governance and Power*. London: Routledge.
Van Maanen, John, Jesper B. Sørensen, and Terence R. Mitchell. 2007. "The Interplay Between Theory and Method." *Academy of Management Review* 32:4, 1145–1154.
Weldes, Jutta. 2006. "High Politics and Low Data: Globalization Discourses and Popular Culture." In *Interpretation and Method: Empirical Research Methods and the Interpretive Turn.* Dvora Yanow and Peregrine Schwartz-Shea, eds. (Armonk, NY: M. E. Sharpe), 176–186.
Wood, Elisabeth Jean. 2009. "Ethnographic Research in the Shadow of Civil War." In *Political Ethnography: What Immersion Contributes to the Study of Politics.* Edward Schatz, ed. (Chicago: University of Chicago Press), 119–142.
Wright, Susan, ed. 1994. *Anthropology of Organizations*. London: Routledge.
Yanow, Dvora. 2000. *Conducting Interpretive Policy Analysis*. Thousand Oaks, CA: Sage.
Yanow, Dvora and Peregrine Schwartz-Shea, eds. 2006. *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*. Armonk, NY: M. E. Sharpe.
Ybema, Sierk, Dvora Yanow, Harry Wels, and Frans Kamsteeg. 2009. *Organizational Ethnography: Studying the Complexities of Every day Life*. London: Sage.
Zirakzadeh, Cyrus Ernesto. 2009. "When Nationalists are not Separatists: Discarding and Recovering Academic Theories while Doing Fieldwork in the Basque Region of Spain." In *Political Ethnography: What Immersion Contributes to the Study of Politics.* Edward Schatz, ed. (Chicago: University of Chicago Press), 97–118.

Ethnographic Innovations in the Study of Post-Communism: Two Examples

Jan Kubik
Rutgers University
kubik@rci.rutgers.edu

In Kubik (2009), I reviewed the three basic types of ethnography and outlined their uses for political scientists. I begin by summarizing my conclusions.

The usefulness of ethnography for comparative politics and political science in general cannot be assessed without realizing that there is no single ethnography. At least three broad types can be isolated: positivistic (realist), interpretive, and post-modern (including multi-sited) ethnography. Each is associated with a different ontology of the social, and each can help political scientists in different tasks.

Positivistic ethnography is indispensable for studying: (a) *overlooked* (informal dimensions of) power; (b) *hidden* (faces of) power; (c) *inaccessible* (mechanisms of) power, for example, in early stages of protest mobilization (d) ostensibly *inconspicuous resistance* to power; (e) *ambiguous* (effects of) power exercise; and (f) *cultural construction* of agents and subjects of power.

Interpretive ethnography is crucial for exposing the relations between power and meaning in concrete situations. Its significance for political analysis has become clearer as a growing number of political scientists—particularly in comparative politics—work within a constructivist paradigm and design their research programs around such principles as: (1) *ontological realism*[1] (focus on actual actions of real people, rather than variables) (2) *constructivism/interpretivism* (focus on "semiotic practices" (Wedeen 2002)), and (3) *micro-focus* on "small scale" mechanisms.

Post-modern ethnography is central for capturing the dynamics of power and identity in an increasingly inter-connected and globalized world. Multi-sited ethnography, attentive to the novel (gradually more virtual) ways of constructing collective identities and focused, *inter alia*, on the increasingly transnational and translocal nature of political and economic transactions, is a promising addition to the methodological armamentarium of today's social science.

Moving beyond my 2009 analysis I want to signal also that the term "ethnography" refers to at least three overlapping yet sufficiently distinct types of intellectual activity and research practice: (1) data collecting, (2) modeling of social reality, and (3) genre of writing. The essence of ethnography as a specific method of data collecting is, of course, *participant observation*. Ethnographic models routinely take form of specifically organized *narratives* built around theoretical claims about the regularities observed in the studied fragments of reality (for example, holism of the social system in early, functionalist, ethnographies).[2] Finally, the narrative, as a type of modeling reality (as opposed to a formal or statistical models), is a *genre of writing* whose tropes can and often are subjected to intense scrutiny, for example with the help of methods borrowed from literary criticism (Clifford and Marcus 1986, for a recent review see Rumsey 2004). The development of the ethnographic method often progressed simultaneously on all three dimensions. Consider Malinowski, who pioneered the method of extensive fieldwork, established the cannons of composing the ethnographic narrative, and had a powerful influence on the literary style of the whole enterprise.[3] Parenthetically, the study of the relationship between models and their dominant *theoretical tropes*, and the *literary tropes* "modelers" utilize, has a long tradition in history (White 1973), sociology, and anthropology (Rumsey 2004); a comparable self-examination is long overdue in political science. In brief, our efforts to find a firmer place for ethnography in political science will become more precise if we remember that each type of ethnography needs to be considered in three dimensions (see Table 1).

The table is intended to signal the complexity of the task that cannot be undertaken in a short essay. Instead of identifying and analyzing examples that would fit all nine cells, I offer only basic ideas and proceed to focus on two brief examples of how ethnographic evidence challenges standard political scientific "findings" in the field of post-communist studies. The first example is intended to show the merits of "positivistic" ethnography (particularly its focus on the informal dimension of organizations); the second illustrates the benefits of ethnographic studies of semiotic practices in small communities.

I am relying on a project I have been working on over the last several years with Myron Aronoff (Aronoff and Kubik, forthcoming). My main task in this project is to understand why political scientists and (cultural, social) anthropologists usually offer different theoretical pictures of what has transpired over the last twenty years of post-communist transformations. By and large, the representatives of each discipline tend to cluster around a relatively predictable set of common assumptions, rely on similar theories, and reach out for similar research tools. As a result, the models or pictures of post-communist "reality" both disciplines offer differ from each other, often quite dramatically. But there are places where these two disciplines enter a dialogue and "correct" each other's models and methods; the consideration of ethnography (in its three dimensions) is one of them.

The differences between these disciplines are ideological, philosophical, conceptual, and theoretical, but they also result from a different calibration of the methodological apparatus: while political scientists routinely rely on "lean" variables sufficiently abstracted from "reality" to allow statistical operations and large-scale comparisons, anthropologists tend to stick to richer, textured descriptions and "within-the-case" analyses of specific instances of a given phenomenon (say, health care reform, privatization, or the change of property regimes). Political scientists recognize the dichotomy as the one between qualitative and quantitative research strategies or between variable oriented and case oriented studies (Ragin 1987). In recent years our understanding of both approaches and their relative (de)merits has grown exponentially (Brady and Collier 2004). The time has come, however, to deepen this understanding by systematically examining ethnography as a specific subspecies of qualitative methods. To illustrate the way in which ethnography modifies, enriches, and sometimes falsifies the models of reality produced by other methods I now turn to two examples.

Is Civil Society Universal?

The first example comes from the studies of civil society. A review of the existing literature on the topic shows that the concept often denotes a specific organizational arrangement in which: (1) a set (or system) of secondary groups (organizations) that are (2) transparent; (3) rely on democratic, deliberative rather than authoritarian style of decision-making; and (4) operate within a public space protected by the rule of law. This specific form of organization may be labeled *legal transparent civil society (LTCS)*.

Furthermore, many political theorists assert the development of civil society (particularly its full-fledged form—*LTCS*) is a significant, if not necessary, component of modernization a lá West. In a nutshell, many observers see the growth of civil society as a part and parcel of westernization-modernization. The argument is that what transpired in the West and contributed to the rise of modern democracy, was the emergence of a specific form of social organization—civil society. This form is characterized by the simultaneous presence of all *four condi-*

Table 1: Three Dimensions and Three Types of Ethnography[4]

		Type of Ethnography		
		Positivist	Interpretive	Post-Modern
Dimension of Ethnography	Fieldwork Method	Participant Observation	Participant Observation of Semiotic Practices	Multi-Sited Participant Observation (Usually of Semiotic Practices)
	Theoretical Trope	(Social) Holism (Social System Composed of Identifiable Segments)	(Semiotic) Holism (More or Less Coherent System of Meanings)	Mosaic
	Literary Trope/ Genre of Writing	Realistic Prose	Intepretive Essay	Collage (Sometimes Poetic) Messy Text

tions outlined above (secondariness, transparency, tolerance, and legality).

According to a related argument, both democracy and civil society can be transplanted into new social and cultural contexts and that the only obstacle is the lack of sufficient resources. Such views, often based on a belief in the possibility of successful institutional engineering *ab nihilo*, are controversial. Some social scientists espouse an optimistic faith in "democratic engineering"; others remain skeptical. Many critical observers (particularly anthropologists) point out that as a product of a specific phase in the development of the Western civilization, civil society may not offer a usable institutional blueprint for non-Western societies.

Others criticize the arrogant ethnocentrism of the assumption that the creation of a more equitable social order and a more democratic polity is impossible without a wholesale borrowing of the quintessentially Western institutional form called "civil society." To combat the dangers of such ethnocentric partiality, some anthropologists and other "anthropologically-minded" social scientists, working mostly in non-Western societies, have attempted to identify a number of institutional arrangements which function as civil society's analogues, yet do not possess all features of the "canonical" Western model. Varshney, for example argues that: "We should look for the alternative civic sites that perform the same role as the more standard civic organizations do" (2003:46).

Researchers engaging participant observation are in the best position to identify, conceptualize, and analyze *actual* institutional and organizational arrangements that emerge both in West and in non-Western sites subjected to various civil society-building or democracy-building projects. In the West, the quality of civil society can be measured on four dimensions identified earlier. The trick is to identify, conceptualize, analyze, and evaluate the non-western equivalents of those four features. Taken together they should amount to social arrangements that: (a) allow people to *organize themselves* "above" the level of kinship; (b) *maintain (a modicum of) transparency* in the public arena; (c) *champion and practice the climate of toleration* in within and between various organizations (including the state); and (d) *constitute at least a tolerated (if not legally sanctioned) counterbalance* to the state's monopolistic tendencies to dominate the public life.

Ethnographic studies provide many examples of such arrangements, functional equivalents of the *LTCS* form. They are sometimes seen as complex hybrids of Western and domestic social forms. Anthropologist David Nugent (1999) studied a hybridization of both democracy and civil society in Latin America. He describes and analyzes the formation of *new politics* (NP) in several Latin American countries, particularly Peru. What is strikingly relevant for my analysis is his demonstration that this innovative form of politics does not reject the classical liberal Western blueprint, but rather it builds on it, adjusting it to the realities of Latin American context. For example, the concept and practice of new politics offer a solution to the problem of *secondariness* in a society organized primarily around primary (or at least non-contractual) groups. Within the Western tradition, civil society (*LTCS*) is construed as a section of the public space formed predominantly if not exclusively out of associations constructed as secondary groups. *Secondariness* is a component of the liberal project: secondary groups are constituted by freely contracting sovereign individuals. Yet as anthropologists have amply demonstrated the sovereign "individual" of the liberal project does not exist everywhere in the world; as a highly specific construct of the personhood it emerged and is dominant only in a subset of "Western" cultures. In a very profound sense, the successful creation of classic civil society (*LTCS*) depends on the supply of specifically "constructed" sovereign individuals and thus on the prior destruction (or at least suspension) of the strong ties that bond people in primary groups. New politics proposes an innovative hybrid as it "accepts" pre-existing collectives (often primary groups) as building blocs of civil society. Varshney (2003) analyzes a structurally similar solution in India where secondary associations provide bridges between Hindu and Muslim communities and in some situation actually contribute to the creation of interethnic peace in the zones of potential conflict.

In short, ethnographic work focused on the *informal* dimension of political and social life seems to be the best research tool to identify *civil-society-like hybrids* that differ from the "Western model" on all four dimensions. Thus, there are hybrid networks that exist somehow between secondary and primary structures. Ekiert and Kubik (1999) identified them in post-totalitarian Eastern Europe; Varshney (2003) analyzed "everyday forms of civic engagement in India." There are semi-transparent institutional solutions: Fox (1994) writes about semi-clientelism in Mexico; Nugent (1999) sees semi-transparency in New Politics in Peru. There are many forms of social organization that are neither purely democratic-deliberative, nor wholly hierarchical. Finally, selective legalization, an imperfect substitute for the rule of law, is a well-studied (often ethnographically) feature of many authoritarian and post-totalitarian regimes and is construed as a key systemic feature that contributes to the liberalization and democratization.

Where is *Homo Sovieticus*?

A critical debate in the field of post-communist studies provides a second example. In this field, scholars ponder a question of why the course of post-1989 transformations has been so uneven, often slow, and certainly costly. Why in some places (Central Asia, Russia, Belarus) it has stalled or careened away into the path of authoritarian reversal? And why, even in the most successful countries, there are persistent "problems," including, for example, underperforming sectors of the economy, unstable party systems, and weak civil societies? Scholars offer many theories, yet in much of what they argue one can detect two broad explanatory logics. For the lack of better terms, I propose a distinction between two (ideal) types of theories: *social adjustment* and *institutional adjustment*. Proponents of both views offer diagnoses (what is wrong?) and policy prescriptions (what is to be done?), but while the champions of *social adjustment* tend to locate the source of troubles in the postcommunist societies, the advocates of *institutional* *adjustment* see the principal source of problems in the reform programs, in their design or implementation, or both. Accordingly, while the former seek to formulate programs of social renewal that are aimed at changing the people so they can "fit" the indispensable institutions, the latter would rather redesign the incoming institutions so they can "fit" the people. A clear articulation of the philosophy of social adjustment comes from the Polish (neo)liberal economist, Jan Winiecki. In an interview he offered the following analysis:

> Q: Who has proved to benefit the least (from the post-communist transformations – JK)?
> A: The problem of Poland is the Poles themselves who wait for a manna from heaven and think that they deserve everything without work and commitment. It is the passive part of society that is at fault. These people are demoralized by the previous system and by those they vote for (in Buchowski 2006).

Piotr Sztompka, an internationally renowned Polish sociologist, advanced an elaborate analysis of cultural mechanisms that contribute to the formation of people "who wait for a manna from heaven." In a 1993 article, he proposed a model of a specific culture, developed under state socialism that has been responsible, at least partially, for turning (some) members of *Homo Sapiens* into *Homo Sovieticus*. Representatives of this sub-specie have difficulties with becoming "proper" participants in a system based on democracy and market economy because they are held down by *civilizational incompetence*. This cultural syndrome is characterized by *seven attributes* that need to be overcome—suggests Sztompka—if the East Europeans want to catch up with the Western part of the continent. He argues that getting rid of the syndrome "is prerequisite, a necessary condition for attaining *true modernity* (original emphasis): authentic democracy, functioning market and open society (1993:91).

I designed an ethnographic test of Sztompka's argument in order to determine if a cultural syndrome he labeled "civilizational incompetence" exists in two regions of Poland I and others have studied using participant observation. Sztompka did not specify where (geographically and sociologically) civilizational incompetence is located thus it is fair to assume that it should be found in *any* region.[5] I interrogated ethnographic data to determine whether, how, and to what degree Sztompka's seven features of civilizational incompetence show up in local/regional (sub)cultures. Or, put differently, I conducted a comparative analysis of two vernacular cultural syndromes shaping people's coping with and engagement in postcommunist transformations (for example, their visions of the political or the normative bases of their trust in authorities). In both cases, the portrayals of regional/local cultures were formed after extensive ethnographic studies (including my own 10-month work in one of the regions). Like Sztompka, I did not focus on *actual actions/behavior* but rather cultural scenarios (Laitin's "points of concern" (1988:589) that serve as *models of or for* actions/behavior (Geertz 1973:93).[6] My project belongs thus to the broad tradition of interpretive ethnography, as I focus on *specific discourses that constitute vernacular knowledge* in each region.

The conclusions from my test are pretty straightforward: (1) there is no evidence of the syndrome of civilizational incompetence in two studied regions. (2) Some individual features of the syndrome can be detected, but they are embedded in region-specific cultural syndromes that have little or no resemblance to the hypothesized cultural entity. (3) Regional (or local) cultures provide people with resources that they utilize in developing strategies of coping with externally imposed cultural, economic, and institutional designs.

Ethnographic studies help to identify and ameliorate the shortcomings of the two broad approaches to post-communism I labeled institutional and social adjustment. Ethnographers and other "ethnographically minded" researchers are guided by the methodological principle of observing actors up close and a theoretical directive of providing a careful and contextualized reconstruction of human agency. If in the view of social adjustors people (often in general) are seen as the "material" that needs to be fixed and while institutional adjustors call for fixing reform programs and their implementation, in the "ethnographic" research program people are seen as agents who are capable of fixing themselves and who, in fact, are always doing so. The proponents of this view emphasize agents' ability to adjust their strategies (we might call their theory: *strategy adjustment*) in order to cope with the changing environment. Additionally, interpretive ethnography allows us to see how cultures are reproduced, sometimes strategically, by specific actions of concrete actors. As (interpretive) ethnographers we have access to localities formed by people who develop common worldviews embedded in practical knowledge. We can, therefore, observe how ensembles of discourses coalesce, climates of opinion solidify, and patterns of choices emerge. We can study how such patterns, that are *ex post facto* detected by public opinion or exit polls, are created, reproduced, and changed within specifically located *trust networks* and complex albeit often *informal power constellations*.

In general, research projects that include an ethnographic component make it possible to observe and reconstruct actors' strategic creativity and thus serve as a welcome corrective to approaches that treat postcommunist transformations as phenomena of the macro-scale, driven exclusively or mainly by mechanisms external to actual locations or communities.

Notes

[1] For the lack of better term I use here "realism," to indicate only that ethnography I am discussing here assumes an ontology according to which social reality is composed of people, their attributes and actions, not variables.

[2] In political science (particularly comparative politics), David Laitin (2004) champions narrative, as one of the three forms of modeling reality. For another take on the role of narrative in political science see Patterson and Monroe (1998)

[3] Malinowski's literary style was clearly influenced by the work of his compatriot, Joseph Conrad (Korzeniowski) (Young 2004). On earlier developments in the history of ethnographic monographs see Thornton (1988).

[4] The table offers very preliminary, imprecise labels. The concept of ethnographic holism is critically discussed in Rumsey (2004). I

take the notion of "messy text" from Marcus (1998).

[5] Mine is however an exploratory study. In future, more comprehensive tests, random sampling of regions is advisable.

[6] Sztompka engages in "the search for underlying *patterns for thinking and doing*, (original emphasis), commonly shared among the members of society, and therefore external and constraining with respect to each individual member" (1993:87).

References

Aronoff, Myron J. and Jan Kubik. Forthcoming. *Anthropology and Political Science: Culture, Politics, and Democratization*.

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

Buchowski, Michal. 2006. "The Specter of Orientalism in Europe: From Exotic Other to Stigmatized Brother," *Anthropological Quarterly* 79:2, 463–482.

Clifford, James and George E. Marcus, eds. 1986. *Writing Culture. The Poetics and Politics of Ethnography*. Berkeley: University of California Press.

Ekiert, Grzegorz and Jan Kubik. 1999. *Rebellious Civil Society: Popular Protest and Democratic Consolidation in Poland, 1989-1993*. Ann Arbor: University of Michigan Press.

Fox, Jonathan. 1994. "The Difficult Transition from Clientelism to Citizenship: Lessons from Mexico." *World Politics* 46:2, 151–184.

Geertz, Clifford. 1973. *The Interpretation of Cultures*. New York: Basic Books.

Kubik, Jan. 2009. "Ethnography of Politics: Foundations, Applications, Prospects." In *Political Ethnography: What Immersion Contributes to the Study of Power*. Edward Schatz, ed. (Chicago: University of Chicago Press), 25–52.

Laitin, David. 1988. "Political Culture and Political Preferences," *American Political Science Review* 82:2, 589–593.

Laitin, David. 2004. "Comparative Politics: The State of the Discipline." In *The State of the Discipline* Vol. 3. Helen Milner and Ira Katznelson, eds. (Washington: American Political Science Association), 147–171.

Marcus, George. 1998. *Ethnography Through Thick and Thin*. Princeton: Princeton University Press.

Nugent, David. 1999. "Democracy, Modernity and the Public Sphere: Latin American Perspectives on North American Models." Prepared for "The Anthropology of Politics and the Politics of Anthropology," Manchester '99. Visions and Voices, University of Manchester, 1999.

Patterson, Molly and Kristen Renwick Monroe. 1998. "Narrative in Political Science." *Annual Review of Political Science* 1, 315–331.

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

Rumsey, Alan. 2004. "Ethnographic Macro-tropes and Anthropological Theory." *Anthropological Theory* 4:3, 267–298.

Sztompka, Piotr. 1993. "Civilizational Incompetence: The Trap of Post-Communist Societies." *Zeitschrift fur Soziologie* 22:2, 85–95.

Thornton, Robert. 1988. "The Rhetoric of Ethnographic Holism." *Cultural Anthropology* 3, 285–303.

Varshney, Ashutosh. 2003. *Ethnic Conflict and Civic Life: Hindus and Muslims in India. Second Edition*. New Haven: Yale University Press.

Wedeen, Lisa. 2002. "Conceptualizing Culture: Possibilities for Political Science." *American Political Science Review* 96:2, 713–728.

White, Hayden. 1973. *Metahistory*. Baltimore: Johns Hopkins University Press.

Young, Michael W. 2004. *Malinowski. Odyssey of an Anthropologist, 1884-1920*. New Haven: Yale University Press.

Shouts and Murmurs: The Ethnographer's Potion

Timothy Pachirat
The New School for Social Research
pachirat@newschool.edu

Breaking News: Political Scientists Announce Invention of Invisibility Potion!

In a stunning disciplinary upheaval unparalleled since the mid-20th century publication of Robert Dahl's "Epitaph for a Monument to a Successful Protest," political scientists spilled from their offices and gathered in campus quads to celebrate today's announcement of the invention of the Fieldwork Invisibility Potion, or FIP. [1] The culmination of decades of top-secret research funded by the Special Operations Branch of the Social Science Directorate of the National Science Foundation, today's release of FIP allows for the first time the possibility of ethnographic field research uncontaminated by observer-observed interactions. Variously termed, "bias" and "subjectivity" by leading political science practitioners, these observer-observed interactions have long plagued the quest for a replicable, objective, and systematic ethnographic method. With FIP's invention, such sources of uncontrolled error in ethnographic research may very well join flat-earth theories, witch burning, and medical bloodletting in the dustbin of prescientific history.

"The observer effect has been a known feature of theoretical physics and quantum mechanics since Heisenberg first articulated it in the early part of the 20th century," remarked an ecstatic Dr. Popper Will Falsify, Principal Investigator on the FIP project. "So it's no small irony that it should be the social sciences that discovered the key to overcoming the Heisenberg effect in studies of the social world. I believe FIP may very well usher in a new era of comity between the so-called social and natural sciences."

Unsurprisingly for a discipline as fraught as political science, however, initial reactions from leading practitioners to the invention of FIP have been anything but unanimous.

Those working within a logic of inquiry informed broadly by positivist commitments are cautiously optimistic about the potential for FIP to add credibility and rigor to fieldwork conducted by political scientists. "Unfortunately," stated Dr. Cy N. Salthaway in an interview conducted at the prestigious Institute for Cumulative Knowledge, "politicians and governments often have a vested interest in portraying certain images of their societies. Ethnography, in-depth immersion, and participant observation are sometimes the only ways of getting a better handle on realities as they actually exist on the ground. But the obvious advantages of immersive fieldwork that gets closer to ground level facts are diluted if not actually reversed by their 'just anecdotal' quality, namely, that because the fieldworker is necessarily only observing interactions highly contingent on her location in time and space, there is really no way to systematize and generalize the data she collects. Of course, another important worry when it comes to researchers who rely extensively or exclusively on ethnographic research is that the data they collect is more an artifact of their presence than a reflection of what is actually there. Combined, these concerns make it exceedingly difficult for ethnography to justify itself as sufficiently systematic or replicable to qualify as science."

As an example, Dr. Salthaway recounted a recent job talk at ICK's Department of Political Science in which the job candidate had invested years to learn Malay and live in a remote Southeast Asian village of approximately seventy families. "There were some remarkably vivid descriptions of pilfering, gossip, and foot-dragging that came out of this fieldwork," noted Dr. Salthaway, "but when a respected senior member of our department interrupted the candidate halfway through his talk to ask whether the research amounted to anything more than an anthropological monograph about a specific researcher living in a specific village at a specific historical moment, it really put a damper on things."

"I'm loathe to throw the baby out with the bathwater," continued Dr. Salthaway, "and it's why I advocate what you might call a three-legged stool approach to the scientific study of politics. Under this approach, ethnography is immensely useful for generating hypotheses, exploring peculiar residuals that appear in statistical analyses, or helping the researcher uncover potential causal mechanisms linking dependent and independent variables. But, ultimately, to produce what I would consider truly valid scientific knowledge, ethnography must be subsumed within a broader research program in which the other two legs of the stool—statistical and formal analysis—serve to test, and ultimately verify or falsify, the hypotheses and hunches developed by fieldwork."

Pressed on whether the invention of FIP changed his basic position, Dr. Salthaway said, "Well, it's an interesting question. On the one hand, by containing the potential to eliminate entirely the participant in participant observation and produce a pure observer qua observer, it does strengthen the capacity of ethnography to be more objective. On the other hand, FIP does not do much for the 'just anecdotal' problem insofar as an observer, no matter how invisible, is still only observing highly specific interactions and settings. So ultimately, I think that even fieldwork conducted using FIP would still need to be combined with statistical and formal analysis."

"I suppose you could say," continued Dr. Salthaway as he leaned slightly forward on the dark red mahogany conference table, "that one anxiety I have is that by making ethnography somewhat more rigorous without overcoming all of its limitations, FIP may give the dangerous illusion of strengthening arguments for the stand-alone value of ethnography in political science. If you'll indulge the extended metaphor, proponents of ethnography's stand-alone value have always seemed to me a bit like creators of one-legged stools. Now, I grant you that one-legged stools might be very aesthetically pleasing, they might make for wonderful conceptual or installation art, and it might even be possible to create an entire tradition or discipline of one-legged stools in which earlier styles are compared with later styles, different types of wood are employed for the stool, different varnishes are put on it, passionate debates erupt over whether this or that wood is more ethical and environmentally sustainable, over whether this or that kind of varnish better respects the underlying grain or "voice" of the wood, and so on and so forth. This kind of thing might continue to the point where these debates replace the actual making of stools as the primary concern of the one-legged stool school or tradition or discipline. But, ultimately, for those concerned with the advancement of *science*, all of this hyper-reflexivity and navel-gazing boils down to the rather straightforward question of whether you would ever want to sit on a one-legged stool. And just as no one would ever want to sit on a one-legged stool, no matter how beautifully crafted, so too would we be better off the sooner we abandoned the fantastical notion that stand-alone ethnography, absent a kind of disciplining or stabilization by statistical and formal analysis, serves to move the project of a scientific study of politics forward. FIP undoubtedly makes the ethnography leg of the three-legged stool of science stronger, but it does nothing to eliminate the need for the other two legs of the stool."

"Now don't get me wrong here," Dr. Salthaway quickly added as he settled comfortably back into his chair. "I have the highest respect for Dr. Popper Will Falsify and what he has accomplished with the invention of FIP, but I just don't think we're quite there yet in terms of a truly scientific, truly replicable, truly systematic ethnographic capacity in political science." Queried about would be needed for such an ethnographic capacity, Dr. Salthaway smiled and said, "Oh, a time machine, for starters. And a do-over button. And a hermetically sealed social world in which our publications had no chance of being read by those they analyze, since that too might alter their behavior." Asked if he was joking, Dr. Salthaway's smile broadened as he chuckled good-naturedly and responded, "If someone had told you a month ago that an invisibility potion for ethnographers would soon be available would you have thought them joking?"

Political scientists working within interpretive logics of inquiry were equally guarded about the invention of FIP, but for very different reasons. Reached via satellite phone while conducting fieldwork on the reinvention of ancient dance traditions as modern forms of resistance to the high modern state among the Eveny reindeer herders of Northern Siberia, it took no small degree of persuading to convince Professor Maura D. Scripshon, founding member of a loose but increasingly influential affiliation of ethnographers in political science known as the Political Ethnography Collective, that FIP was not simply an elaborate hoax.

"Like all revolutions in technologies of observation and analysis from the microscope to the telescope to the explication of the bell curve to the development of ordinary least squares analysis to the delineation of fuzzy set analysis, this so-called FIP will no doubt be heralded by many as a breakthrough of magnificent proportions and magical possibilities," said an audibly skeptical Dr. Scripshon. "Fundamentally, however, I do not think the invention of FIP or any other optic technology obviates some of the basic, unavoidable questions facing the ethnographer and, by extension, all who take the social world as their focus of analysis. Indeed, a central motivation for organizing PEC, which now has close to forty active members and just celebrated its first anniversary last week, is the contention that the power of ethnography lies not only or even primarily in its capacity to get closer to the ground, to better 'collect data' as if data are like so many rocks lying about in a field, but rather precisely in the way ethnography forces us to confront the question of how we as researchers are implicated in the social worlds we study, to confront the ways we actually co-generate rather than simply collect data, and to confront the ways the knowledge we produce with this data travels back and alters the very social worlds it purports to explain."

After excusing herself briefly to attend to a reindeer making strange clicking noises that interfered with the phone conversation, Dr. Scripshon continued, "The idea of neutrality or objectivity in fieldwork is an illusion because the participant observer is always intervening in specific relations and networks of power. Take as one example a researcher who studies social relations of production on a factory floor. Not only *what* but *how* that researcher sees is going to be intimately tied to whether or not she enters the factory as a guest of management or whether she enters as an entry-level line worker, just to contrast two starkly different positional embodiments the researcher might take. Further, the ethnographer is always situated at the intersection of multiple identities—racial, gendered, sexual, class, and so forth—and these impact not only how people in the field interpret and therefore respond to her but also how she herself filters her observations. So arguably, the more fraught the power relations in the field, the more accounting for these kinds of positionality matters to the quality of the research."

"Forgive the Russia-inspired metaphor," shouted Dr. Scripshon in an effort to be heard as the clicking reindeer noises grew into a concerted chorus of bellowing, "but this kind of researcher-specific positionality is really just the most obvious and least avoidable center of a successive series of matryoshka dolls. Ethnography's attention to researcher-specific positionality is nested inside another ontological and epistemological doll that asks the researcher to explicitly account for the ways in which the underlying logic of inquiry used in the research channels a whole series of decisions of great import beginning with the framing of the research question to the way the researcher counts certain things as facts or observations relevant to the research and others as coincidental or unimportant. And this second ontological doll is itself nested within yet another, third doll that asks the researcher to locate her research project in relation to a larger disciplinary history that is connected to broader political projects, funding programs, and specific ideologies and interests. Empires deploy armies of scholars as well as armies of fighters, and the third level of this matryoshka doll asks the researcher to give an accounting of the *uses* of research, of the kinds of discourses one's research legitimizes and is in turn legitimized by, and of the likely effects—intended and unintended—of those discourses on the subjects of research and the broader social and political worlds they inhabit."

Pressed on whether the invention of FIP altered her matryoshka doll approach to ethnography, Dr. Scripshon responded, "The strength of ethnography, even for those who will never use ethnographic research methods, is that as a method it is especially suited to surfacing troubling and important questions about positionality and power at the level of the researcher, the level of underlying logics of inquiry, and at the level of the discipline itself. Arguably, all research methods in political science—methods ranging from immersive ethnography to elite interviewing to survey research to focus groups to regression analysis to formal modeling—implicitly or explicitly provide answers or a range of answers to the questions implied in a matryoshka doll approach. But because ethnography posits the embodied researcher as the instrument of research, it is unique as a method in inviting reflection on what is often silenced, controlled for, or completely neglected in other methods, namely, the central role of the researcher in co-generating knowledge about the social world and her positionality within unequal networks of power in co-generating that knowledge. What, after all, are the practices of research, the practices of producing knowledge and truth claims about the social world and publishing those claims in specialized legitimating forums, if not acts of intervention and meaning-making? Ethnography, in this sense, makes visible the degree to which the image of the objective social scientist dispassionately removed from the worlds she studies is nothing more than a convenient—if sometimes productive and enabling—fiction, a fiction that both produces and reproduces certain power effects in the social worlds it studies. And so the accompanying worry for me, when it comes to FIP, is that rather than rendering such questions about power and positionality irrelevant, FIP simply makes them even less likely to be surfaced within the discipline of political science by furthering what can ultimately only ever be an *illusion* of invisibility."

Asked whether she would consider using FIP during her fieldwork amongst the Eveny, Dr. Scripshon laughed and said, "Do you hear those reindeer in the background? I really must go feed them before they start a stampede. But in response to your question, let me just ask if after three years of slowly building rapport with this community and countless bruises and not a few broken bones from trying to master the impossible art of riding smelly reindeer across the tundra, you really think I would squander all that hard work by swallowing some potion that makes me disappear with a puff and a poof?"

Drs. Salthaway and Scripshon's contrasting reactions to FIP suggest that disagreements over the place and value of ethnographic research in political science are rooted in longstanding disciplinary debates defying easy resolution. Meanwhile, political science departments and Institutional Review Boards are scrambling to deal with the practical implications of FIP's imminent release. In an effort to attract the most competitive Ph.D. applicants, some top departments are already promising funding and specialized methods courses to support FIP-enabled fieldwork while other departments are embroiled in debates over whether FIP ought to be reserved for tenured faculty, at least in its initial years of use. And in keeping with the patchwork system of university-specific Institutional Review Board procedures, some IRB committees are all but requiring ethnographers to use FIP in the field, arguing that harm to subjects is radically reduced by the invisibility of the researcher, while other IRB committees have taken an opposite approach, equating invisibility with a kind of deception which requires a special and difficult to obtain exemption from standard informed consent requirements.

For Principal Investigator Dr. Popper Will Falsify, the intellectual and pragmatic debates ignited by the invention of FIP only serve to underscore its revolutionary importance for political science. "We may very well be standing at a historic junction," he stated confidently. "In one hundred years, the history of political science may be divided simply into pre-FIP and post-FIP. Real progress has been made. There will be no turning back."

Note

[1] This essay is a playful thought experiment. For their insights, I thank, without implicating, Lisa Bjorkman, Lee Ann Fujii (who, among other things, came up with names for Drs. Salthaway and Scripshon), Clarissa Hayward, Courtney Jung, Brandon Koenig, Jan Kubik, Dorian Warren, Lisa Wedeen, Emily Wills, and Dvora Yanow. And, a special thanks to Ed Schatz for organizing the 2009 Toronto APSA roundtable on Ethnography in Political Science and for editing this symposium.

Studying Politics with an Ethnographic and Historical Sensibility

Dorian Warren
Columbia University
dw2288@columbia.edu

What are the implications of studying political phenomena where we don't know the outcome?[1] How might we think through the differences in studying political events, interactions, processes and outcomes in "real-time," compared to historical approaches, where we study important events that have already happened (whether social movements, political change, etc.)? I come at these questions wearing two methodological hats within American politics: one as a political ethnographer, and the other as a toiler in the historically-oriented subfield of American political development. My use of political ethnography as an approach primarily rests on neo-positivist and realist ontological assumptions (Schatz 2009; Kubik 2009; Allina-Pisano 2009; Shapiro 2005). It is from these assumptions that I argue there are consequential differences when studying events or processes in hindsight versus real-time. I draw on my own fieldwork on diversity in American labor unions, union organizational change, and anti-Wal-Mart political campaigns to examine these issues.

Historical Approaches Meet Ethnography

While Ed Schatz reminds us that political ethnography does not exist in a methodological vacuum vis-à-vis the discipline of political science, there has been little explicit engagement of the recent revival of ethnography with the "historical turn" in political science (McDonald 1996). Responding to the rise and dominance of positivism and the post-war behavioral revolution of the 1950s and 1960s, the late 1970s ushered in a resurgence of history and historical approaches to big questions in the American social sciences (McDonald 1996; Sewell 1996; Steinmo et al. 1992; Orren and Skowronek 2004; Mahoney and Rueschemeyer 2003). Scholars seeking clues to the causes and consequences of social revolutions, the emergence and development of social welfare states, or the causes and meanings of political change turned to historical research and historiography to gain a different type of causal leverage unique from quantitative behavioral assumptions (Brady and Collier 2004). In sociology, this became institutionalized as comparative-historical research and political sociology, while in political science it took the form of historical institutionalism in dual and often bifurcated comparative and American versions.[2]

In his methodological primer on the discipline and practice of history, Chicago-School historian Louis Gottschalk (1969) discusses the inherent limitations facing historians and historical research: "Most human affairs happen without leaving vestiges or records of any kind behind them. The past, having happened, has perished forever without occasional traces…only a small part of what happened in the past is ever observed" (45). He continues by discussing the vast majority of "unobserved happenings" in the world,

> And only a part of what was observed in the past was remembered by those who observed it; only a part of what was remembered was recorded; only a part of what was recorded has survived; only a part of what has survived has come to the historians' attention; only a part of what is credible has been grasped; and only a part of what has been grasped can be expounded or narrated by the historian. (45)

Gottschalk goes on to make a distinction between "history-as-actuality" and "history-as-record", arguing that history can only be told from "history-as-record," or the "surviving records of what happened" (46). But ethnographers in this sense are also historians; we document, from our own and from actors' experiences whom we observe, history-as-actuality. An ethnographic standpoint allows the recording of a history in real-time, the making of history, as it were.

If true, then ethnographers perhaps engage in real-time historiography.[3] As observers, we write a history as political actors make it. Instead of reconstructing a past, we "imaginatively construct" the present. The ethnographer is herself a primary source of her data, while also documenting the testimonies of other primary sources.[4] Indeed, the first of Gottschalk's four general rules for significance given to historical documents is concerned with temporality, albeit from a neo-positivist perspective:

> Because a witness's reliability is, in general, inversely proportional to the *time-lapse* between the observation of the event and the witness' recollection, the closer the time of making a document was to the event it records, the better it is likely to be for historical purposes (90).

When conducting research with a much longer time-lapse between observation and recollection, Jacob Hacker and Paul Pierson (2002) warn historically-oriented scholars about the dangers of "post-hoc correlations," or reading back actors' preferences from policy outcomes. There is always the possibility, they point out, that "congruence between preferences and outcomes may be brought about by policies themselves, rather than the enactment of policies reflecting preferences" (285). Political outcomes and consequences of events look more inevitable in the rearview mirror than they often are or were. And political actors' interpretations of the past are often instrumental; they are streamlined and more self-serving towards some political end rather than messy, indeterminate or convoluted. Ethnography addresses this issue in that it forces the researcher to deal centrally with uncertainty, contingency, and actors' interpretations of events as they unfold.

For example, ethnographic fieldwork for my first project focused on the question of diversity and organizational change in the labor union context (Warren 2005). I wanted to understand how a local union of mostly female, black and immigrant hotel workers was transformed from a racist, undemocratic, conservative and moribund organization into a progressive, democratic and revitalized one. More than two years of participant-observation in the midst of this transformation enabled me to examine the internal processes and interactions between workers, union staff, and elected leaders; between workers, union staff and hotel managers; and between workers, union staff, elected leaders, and "outside" allies of the union. I argue that one aspect of organizational change and the inclusion of marginalized workers can be seen through changes in the unions' "identity practices" (Warren 2005). These include union demands vis-à-vis employers and the political system, the internal "culture" of the union, and the collective identity of the union. These identity practices of demands, culture and collective identity were reconstituted and remade through intensive, deliberate and repeated attempts at meaning-making by union staff and workers about what a "union" is, does, and could be.

The main point I want to make here is about one of the punchlines in my analysis. Contrary to romanticized notions of workers of the world rising up and uniting in unbreakable solidarity, this process of organizational transformation was not a democratic one. Far from spontaneous and democratic rebellion, organizational change was initiated and led by those at the top and with centralized power in the union: the union *staff*. And although the outcome as we can see it today is a radically democratic union, the catalyst and process of getting there was far from it. What is important about this observation and argument is that if I were asking the same question about union transformation today, that is, from a historical perspective (as the changes took place from roughly 2000-2003), what I would "see" or "observe," and indeed be told in interviews with workers and staff, would be a fundamentally different causal narrative about organizational change. It would consist of a streamlined story about how Chicago's black and immigrant, mostly female hotel workers became fed up with the inequality they were experiencing in their lives as a result of globalization and the low-wage hotel industry, how they decided to stand up to the their old union leaders and the big hotel companies, and how they ultimately changed their union and won several major fights against the industry.

How do I know this? Because this is the story the union, including workers and staff, *tell themselves*, repeatedly. They do it through the ongoing practices of meaning-making and production of collective identity via print publications (newsletters, flyers, etc.), through oral histories, through organizational practices and through collective action. I have heard this through subsequent interviews with workers and staff members, many years removed from the actual transformative events. If I went back now and accessed all of the union's internal newsletters and other written documents from the last ten years, without an ethnographic sensibility (Pader 2006) or without having documented the changes in 'real time' via observation and interviews, this radically different narrative is what I would find. And on a couple of "revisits" two and three years after having exited the field, in talking with workers and staff, this is also the narrative I have been told. What I would have missed from using a strictly historical approach was not only a different "truth" or causal story (Allina-Pisano 2009), but, more consequential in my view, I would have missed a crucial operation of power within an organization of the "powerless": the production of political memory and collective identity within an organization.[5]

Problems of History and the Promise of Political Ethnography

If the endemic problem of time-lapse wasn't enough of a challenge for neo-positivist historical work, *historical social science* has its own additional baggage, suffering from the consequences of what William Sewell (1995) calls "experimental temporality": multiple occurrences of an event are assumed to be both equivalent and independent. In discussing Theda Skocpol's classic *States and Social Revolutions*, Sewell argues that her work requires "unhistorical assumptions about temporality," reflected in Skocpol assuming her three great revolutions are interchangeable as far as causal conditions are concerned, and are independent "trials," that is, early revolutions could not have had any influence on subsequent ones (258–259).

Ethnography guards against this fallacy. In interviews with twenty-five political actors (activists and politicians) leading the anti-Wal-Mart effort in Chicago during research on opposition to Wal-Mart stores in black communities in L.A. and Chicago, for instance, not one of them mentioned the anti-Wal-Mart campaign that took place a year earlier in Los Angeles, California. Neither was any mention of the successful effort to oppose Wal-Mart in L.A. mentioned in any documents, or oral or written records of the conflict in Chicago. Had I been doing historical research and constructing a narrative based on documents and interviews, I would have had two, equivalent and unrelated episodes of anti-Wal-Mart conflicts. But alas (for such research narratives), there are an ethnographer's fieldnotes! And upon re-reading my fieldnotes from research two years prior to my interviews, I found my transcript of a meeting of anti-Wal-mart coalition activists where several report on discussions held with activists in L.A. Not only was there an exchange of strategy (several L.A. activists told the Chicago activists their interpretation of what led to their success against Wal-Mart), but the L.A. people also sent the Chicago activists, via email, copies of flyers and propaganda used in their campaign (Warren 2009).

These political actors, whether intentionally or unintentionally, had "forgotten" an explicit episode of organizational learning from other actors. One possible explanation of this lost memory is that as the campaign developed into what I call "Act 2" (there are currently four acts in my narrative) and took on a significantly different character, collective memory and learning that had occurred were no longer seen as relevant, and then were forgotten altogether. Another possibility is that as the activists embarked in Acts 3 and 4 on a radically different and novel strategic route, and as this strategy resulted in what became a temporary victory, the narrative that my informants told me *after the events unfolded* and told themselves required the *purposive forgetting* of any advice or strategy shared with them by earlier campaigns (Warren 2009)

What is significant about this is that my own analysis of these campaigns would have fallen into the "experimental temporality" trap if not for ethnography. These campaigns were not independent events; one group of actors learned from their interaction with another. As Sewell (1995) notes with respect to Skocpol's three revolutions, the Russian revolution was in many ways self-consciously patterned on the French revolution, while the Chinese revolutionaries modeled their actions and received direct assistance from the Bolsheviks. Thus, Skocpol "freezes" and "fractures" history "by treating the histories of the three revolutions as if they took place in isolation from one another rather than as a sequence of historically connected events" (Sewell 1995: 259). Anti-Wal-Mart campaigns are far from revolutions, but the lesson applies nonetheless.

What a political ethnographic approach allows us to capture is Sewell's conception of the "radical contingency" he contends is "fundamental to an eventful temporality" (263). But what a political ethnographic approach with a historical sensibility allows is two things that Sewell (1995) brings to our attention: first, it allows us to capture how open-endedness and uncertainty, contingency, and agents' actions "make transformative events possible in the first place" (271). Second, it allows us to capture how those events or actions then come to be "retrospectively appropriated to institute and reproduce structures" (271). In other words, we can study actors and actions in "real-time," actors' own interpretation of those events, and then subsequent reinterpretations of those interpretations. This opens up lines of inquiry into overlooked phenomena such as the production of political memory and the processes of collective identity formation and reproduction within groups, organizations or communities.

These observations implore scholars of American political development, historical sociology, and organizational change to learn the lesson that studying processes and events ethnographically in real-time teaches us: that political, organizational or institutional change often has no inherent "logic" to it; that instead, some element of change is almost always an ambiguous and unintended/unanticipated outcome.

Outside of participant-observation, the closest we can get to an understanding of actors' intentions and self-understandings in their own historical moments is through what Diane Vaughan calls "historical ethnography," itself drawn from Michael Burawoy's (2003) notion of an "archeological revisit." In historical ethnography, the researcher attempts to "elicit structure and culture from the documents created prior to an event in order to understand how people in another time and place made sense of things" (Vaughan 2004: 321). In this way, I believe that this might be doing history with an "ethnographic sensibility" (Pader 2006), while those of us doing ethnography should employ a "historical sensibility."

If one of the major problems with historical social science is how it conceptualizes temporality, then for some questions or puzzles, political ethnography is a well-poised remedy. Another road, then, to an "eventful temporality," as William Sewell urges, is through political historical ethnography. The ethnographic path guards against a teleological and experimental temporality, and emphasizes an "eventful" one through real-time observation and analysis. In a real-time setting, where outcomes, events and processes are open-ended and unknown, teleological explanations are put under severe strain.[6] An ethnographic sensibility instead starts from an open assumption about processes, contingency, and actors' agency. Thus, one is less likely to attribute some "cause of a historical happening…to abstract historical processes leading to some future historical state" (Sewell: 247).

This discussion leads me to emphasize a key point: there are consequential differences when studying events or processes in hindsight versus real-time. For contrary to the common refrain, hindsight is never 20/20. Both actors and researchers are fallible when attempting to tell the story of the past in the present. If, as Vaughan (2004) tells us, "knowing of some harmful outcome, the tendency [of actors] is to focus in retrospect on all the bad decisions that led to it" (333), then the opposite is also true: knowing of some positive outcome (a political victory or momentous event), actors are likely to focus in retrospect on all the good decisions that led to it, and especially to emphasize *their* decisions and actions which they might interpret as especially important. Observing politics in real-time guards against this, as does a historical sensibility.

But one challenge of political ethnography when thinking with a historical sensibility is not only "being here and being there" in terms of two different communities, as Geertz pointed out (1988). The challenge is to "be here and now" and also to be "there and then" (Katz 2004). This is also the method of the foremost ethnographer of American politics, Richard Fenno (1986), describing a politician's "career sequence" at multiple points in time: "whenever you observe a politician, he or she is at some stage in a career that stretches back in time and reaches forward in time. His behavior can be interpreted from the perspective of a career path that brought him to where he is, and of the career path he expects will take him where he wants to go" (11). Indeed, especially in the case of what we might think of as ambitious political or organizational actors, understanding one's long-term ambitions and aspirations, as well as their current understanding of their "reputations," can help us interpret their short-term political interpretations and actions (Fenno 1986).

This leads to an essential tension between a presentist bias on the one hand and historical fallacies on the other.[7] But a political/historical ethnographic approach might help us resolve such a tension. What happens, via political ethnography, when we study processes, events, outcomes in real-time? Uncertainty, improvisation rule the day. Unexpected and unanticipated things occur, which actors interpret, in real time, on the fly, and then often restrategize and act on. There is never a moment of complete information. Life is open-ended; nothing is or looks inevitable. There is no easily identifiable critical juncture. But this then also opens the possibility that how actors interpret actions, events and outcomes often change; the meanings imputed to events or processes may change over time, and if so, how do we deal with this? We can explain political actors' understandings and interpretations of any given moment in real-time, as well as access their own narrative of the same event or moment from their historical perspective.

In his preface to *The Making of the English Working Class*, E.P. Thompson is poignant on the issue of understanding the reproduction of class relations that is relevant for our discussion: "If I have shown insufficient understanding of the methodological preoccupations of certain sociologists, nevertheless I hope this book will be seen as a contribution to the understanding of class. For I am convinced that we cannot understand class unless we see it as a social and cultural formation arising from processes *which can only be studied as they work themselves out over a considerable historical period*" (emphasis added). The implication I take here from Thompson is that in the study of relations of inequality, power and politics, we must be both deeply ethnographic and deeply historical.

Notes

[1] Thanks to Ed Schatz for coordinating this effort. Specials thanks to Dvora Yanow for inspiring and commenting on this essay and Tim Pachirat for his incisive comments.

[2] The comparative politics subfield focused on comparative political economy and welfare state development, while in American politics this took the form of the institutionalization of the separate subfield of "American political development."

[3] Gottschalk defines "historical method" as "the process of critically examining and analyzing the records and survivals of the past,"and "historiography" as "the imaginative reconstruction of the past from the data derived by that [historical method] process" (52-53).

[4] Gottschalk defines a primary source as "the testimony of an eyewitness…present at the events of which he or it tells" (53).

[5] This observation of another, more consequential, question about the operation of power in an organizational context provides potentially fruitful empirical ground for both neo-positivist and interpretivist approaches to political ethnography, *within the same project*.

[6] This is not to imply that a Marxist ethnographer, for instance, might still operate under teleological assumptions while in the field, which would obviously influence the writing of field notes and narrative.

[7] Though a study of "historical myth-making" in itself would prove quite interesting.

References

Allina-Pisano, Jessica. 2009. "How to Tell an Axe Murderer: An Essay on Ethnography, Truth, and Lies." In *Political Ethnography: What Immersion Contributes to the Study of Power*. Edward Schatz, ed. (Chicago: University of Chicago Press), 53–73.

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

Burawoy, Michael. 2003. "Revisits: An Outline of a Theory of Reflexive Ethnography." *American Sociological Review* 68:5, 645–679.

Fenno, Richard F., Jr. 1986. "Observation, Context, and Sequence in the Study of Politics." *American Political Science Review* 80:1, 3–15.

Geertz, Clifford. 1988. *Works and Lives: The Anthropologist as Author*. Palo Alto: Stanford University Press.

Gottschalk, Louis. 1969. *Understanding History: A Primer on Historical Method*, 2nd ed. New York: Knopf.

Hacker, Jacob S. and Paul Pierson. 2002. "Business Power and Social Policy: Employers and the Formation of the American Welfare State." *Politics & Society* 30:2, 277–325.

Katz, Jack. 2004. "On the Rhetoric and Politics of Ethnographic Methodology." *Annals of the American Academy of Political and Social Sciences* 595, 280–308.

Kubik, Jan. 2009. "Ethnography of Politics: Foundations, Applications, Prospects." In *Political Ethnography: What Immersion Contributes to the Study of Power*. Edward Schatz, ed. (Chicago: University of Chicago Press), 25–52.

Kurtz, Sharon. 2002. *Workplace Justice: Organizing Multi-Identity Movements*. Minneapolis: University of Minnesota Press.

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

McDonald, Terrence J., ed. 1996. *The Historic Turn in the Human Sciences*. Ann Arbor: University of Michigan Press.

Orren, Karen and Steven Skowronek. 2004. *The Search for American Political Development*. New York: Cambridge University Press.

Pader, Ellen. 2006. "Seeing with an Ethnographic Sensibility: Explorations Beneath the Surface of Public Policies." In *Interpretation and Method in Political Science*. Dvora Yanow and Peregrine Schwartz-Shea, eds. (New York: M.E. Sharpe), 161–175.

Schatz, Edward, ed. 2009. *Political Ethnography: What Immersion Contributes to the Study of Power*. Chicago: University of Chicago Press.

Sewell, William H., Jr. 1996. "Three Temporalities: Toward an Eventful Sociology." In *The Historic Turn in the Human Sciences*. Terrence J. McDonald, ed. (Ann Arbor: University of Michigan Press), 245–280.

Shapiro, Ian. 2005. *The Flight From Reality in the Human Sciences*. Princeton: Princeton University Press.

Steinmo, Sven, Kathleen Thelen and Frank Longsreth, eds. 1992. *Structuring Politics: Historical Institutionalism in Comparative*

Analysis. New York: Cambridge University Press.
Thompson, E. P. 1966. *The Making of the English Working Class*. New York: Vintage.
Vaughan, Diane. 2004. "Theorizing Disaster: Analogy, Historical Ethnography, and the Challenger Accident." *Ethnography* 5:3, 312–347.
Warren, Dorian. 2005. *A New Labor Movement for a New Century? The Incorporation of Marginalized Workers in U.S. Unions*. Ph.D. Dissertation, Yale University.
Warren, Dorian. 2009. "Contesting Neoliberalism in the City: Grassroots Mobilization and the Political Geography of Anti-Wal-Mart Campaigns." Paper presented at the Annual Meeting of the Southern Political Science Association.

Ethnography and American Political Science: Two Tribes, Briefly Characterized

Edward Schatz
University of Toronto, Canada
ed.schatz@utoronto.ca

Of all the "tribes" that together constitute academic political science, this brief essay considers two.[1] I hope to show that we can learn much about the sociology of academic political science by (artificially) limiting discussion to these two.[2]

The first is a group of tribespeople from the well-known but little-studied region of APSR-stan. The second is a semi-nomadic tribe of political ethnographers. Each has its own rituals, quirks, and forms of boundary maintenance, internal contestation, power, and diversity. One could imagine a full study of the dynamics of each, but let me limit myself here to noting a few of each tribe's "assumed givens" (Geertz 1973: 259). My purpose is not necessarily to call tribal assumptions into question, but rather—in the spirit of Spiro's (1990) invocation of T.S. Eliot—to "make the strange familiar and the familiar strange" as a way to suggest what each stands to gain from engagement with the other.

Assumptions that Prevail in APSR-stan

First consider the people of APSR-stan. The primary distinguishing feature of this tribe—more than any other—is a marked tendency to assume *a priori* the universality of the tribe's propositions, perspectives, and experiences. This is clear if we examine the language that tribespeople use in their premier publication, the *American Political Science Review*.[3] When this journal publishes a single-country study, the presentation of findings depends on what country is being examined. If the single country is the United States, APSR authors overwhelmingly use "unmarked" forms: "Senators and their Constituents" or "What We know about Voting Behavior." These hypothetical titles—I do not here need to single out any particular APSR article to make the general point—do not put into the foreground the temporal or geographic location from which evidence comes; the experience of the United States is assumed to be universal. While the authors of such articles may be open to being proven wrong, the language used shifts the burden of proof: it is up to the critical reader to demonstrate that the US experience is unusual (even if the conclusions advanced are ultimately based on evidence adduced from interviews conducted in Greeley, Colorado in the fall of 1984).

By contrast, APSR authors have tended, over the vast span of the journal's existence, to "mark" evidence that comes from other single-single country contexts. Thus, an article is not simply about voting behavior but about voting behavior with "evidence from Brazil." A piece is not about legislative committees but about committees in the "Israeli Knesset." The normality of the US experience is claimed, the abnormality of other experiences asserted—something achieved through the linguistic practices of marking and non-marking.

That these assumptions are problematic seems plain, especially given what we know about the exceptional aspects of American political development, but what then allows the axioms to be so persistent? After all, the people of APSR-stan are an intensely smart lot, so what keeps them from turning their critical faculties on themselves and the way they practice their craft?

I suspect that two factors are crucial. First, the residents of APSR-stan are overwhelmingly of American stock. (By this, I mean that they are natives of the United States.) Yes, Americans are diverse; no, Americans do not speak with one voice. But the fact remains that APSR-stanis tend to encounter and study non-Americans and their settings quite rarely, in part because the number of American political scientists is so large relative to that of other national communities. As a result, APSR-stanis can publish (as this is what tribespeople in this region value above all else; see Luke [1999]) and thrive almost entirely without attention to or concern for what other, non-American tribes and their academic families are up to. The assumed normality of the American experience as a matter of course is hardly ever challenged.

There is a second reason why APSR-stanis tend to assume the normality of their experience. With rare exceptions, APSR-stanis believe that striving to generate objective knowledge is a virtuous vocation; "objectivity" is this tribe's main currency. And why not? After all, who would argue that they instead should become partisans in political battles? Whatever the objective value of this search for objectivity, think for a moment about the sociological consequences for the tribe; it becomes a matter of normal professional "presentation of self" (Goffman 1959) to put into the background those aspects of one's intellectual work and one's research design that might be perceived as failing to meet the "objective" ideal.

There is, therefore, in this tribe a seemingly strange, and yet—upon some reflection—wholly explicable synergy: in a macro-environment where exceptionalism mas-querades as universalism, the same pattern is replicated at the micro-level. Individual researchers—clearly exceptional human beings in any number of ways—tend to assume the universality of their vantage-points, their observations, their instruments, and their interpretations. Just as the American experience is assumed to be unremarkable, the experience of an individual researcher who strives for objectivity is likewise assumed to be unworthy of note. All of this squares with the tribe's prevailing notion of science as the technical application of universally available tools to research problems. Both the uniqueness of American-dominated perspectives in general and the uniqueness of individual research perspectives in particular are simply rendered invisible.

To be sure, there is more going on in APSR-stan than I can adequately capture. For example, political theorists' work routinely appear in the pages of the APSR—a fact that APSR-stanis treat as evidence of diversity. In addition, the APSR underwent nearly a decade of tumultuous internal change during much of the 2000s, with some tribal traditions apparently attenuating. Recent developments in the journal, however, suggest that reports of the death of tradition are greatly exaggerated.

Assumptions that Prevail Among Political Ethnographers

A second tribe—one of political ethnographers—resides nearby. Unlike APSR-stanis, who are defined by place of publication and the two-tiered assumption of universalism, the members of the political ethnography (PE) tribe are defined by common method, by a fiercely guarded sense of their own individualism, and by a penchant for problematizing.

I should clarify that the members of the PE tribe might not even think of themselves as a group. Indeed, to the extent that political ethnographers have called into question the project of a cumulative and basically progressive science of politics, their endeavor does not automatically lend itself to community building. If APSR-stanis understand themselves as bound by a commonweal project of cumulative knowledge, the PE tribe consists of hardscrabble individuals who may revel in nonconformity. But, like acephalous societies everywhere, this one builds informal institutions to protect against the community's dissolution in the face of centripetal forces; there is, in this sense, still a *group* here.

Second, political ethnographers assume that the outskirts of their larger professional communities are the place to be. Outskirts provide a wonderful vantage-point for seeing and making sense of the rituals and knowledge-claims of academic political science, much as they do when they conduct their ethnographic research. From their metaphorical perch, the strangeness of many of the central features of the mainstream community becomes clear. The practices that dominate in APSR-stan look particularly strange. And yet, as sure as they are that the prevailing orthodoxy is problematic, they are an uneasy lot: they would feel deeply uncomfortable about the construction of a new orthodoxy—ethnographically inflected or otherwise—whose behaviors and rituals might prove to be equally troublesome.

As a result, tribe-members both suffer from, and yet thrive in, a crisis of identity and authenticity. On the one hand, they revel in their marginal status, certain that they are performing a service for professional political science by keeping the mainstream in check and knowing that they have glimpsed truths that are inadequately captured via other approaches. On the other hand, they remain strangely vague about what an ideal future for the discipline would be.

Third, political ethnographers on the whole harbor a preference for critiquing received wisdom. Their descriptions are not merely "thicker" than those offered by other tribes; descriptions are offered in order to call into question established thinking. (The irony, of course, is that their own work simply could not exist without an identifiable and problematic mainstream to critique.) And so, the research questions this tribe asks tend to complicate simple narratives: to unpack, to unbundle, to unsettle, and to destabilize.

There may be good reasons to critique received wisdom, as Calvin Chen (forthcoming) reminds us in describing how his Chinese interlocutors deliberately deceived pollsters. My point is again a sociological one: this tribe, like others, is comfortable asking certain kinds of questions rather than others. In fact, the questions asked tend to serve as a form of boundary maintenance for the group.

What do I mean? When James Scott (1985) highlights the "weapons of the weak" such as "foot dragging," feigned ignorance, gossip, and "character assassination," it is hard to imagine other ways to shed light on these politically significant phenomena. But if ethnography is indispensible—simply and unambiguously necessary—to studying such phenomena, this serves to keep those who lack ethnographic skills or training from considering them directly. Thus, "weapons of the weak" become what ethnographers study (a sort of protected epistemic space), while other phenomena are what non-ethnographers examine. This may not be particularly troubling for those who derive their sense of self from cultivating their individuality and distinctiveness, but it does tend to shut off conversation with outside tribes. Whatever the intention of PE tribespeople, such practices tend ultimately to preserve the boundaries of their community.[4] They are not unique in this sense; other epistemic communities do the same, but this characteristic is at least as notable among the PE tribe as it is elsewhere.

Toward Forms of Exchange

These are some of the basic "assumed givens" of these two tribes, who rarely encounter one another. Something interesting might result, however, from their more frequent meeting. It might even be mutually enriching. In a hypothetical encounter between the ethnographer tribesperson and the APSR-stani tribesperson, the former might ask the latter to clarify the rationale underlying the analytic categories routinely assumed to be universal. The latter might offer a persuasive response or she might refine her use of categories after the encounter, newly self-conscious about her assumptions. In turn, this decentering of American political science might go some ways toward decentering the notion of how social science is conducted. What was once assumed to be universal (the "view from nowhere," in Nagel's [1989] phrasing) might come to be seen as rather more rooted in particular encounters of specific researchers with specific individuals being researched.

At the same time, the APSR-stani might concede that asking unique questions and studying hard-to-reach phenomena brings real value, but nonetheless invite the PE tribesperson to consider some of the same questions that non-ethnographers ask. This would be a profound intellectual challenge—to address questions asked by non-ethnographers in ethnographic ways—but it would generate a variety of interesting substantive conversations.

We have examples of such work. Wedeen (1999) uses ethnography to ask questions about compliance and participation under authoritarianism. These questions are part of the same broad *problématique* considered by other scholars of authoritarianism. Walsh (2004) uses participant observation to study public opinion formation—topics that build upon and engage questions about public opinion asked by non-ethnographers. Arias (2009) argues that we should study abiding problems about Latin America—governance, violence, and statehood—via ethnographic approaches. Similarly, we would learn much about the financial crisis of 2008-9 by examining cultural practices and personal networks among financial planners, hedge fund managers, and bankers, much in the way that Ho (2009) does for Wall Street investment bankers. Such a study could illuminate reasons for the recent unraveling of trust in the US and global capitalist markets.

Some tribespeople from each group may eventually intermarry. Others may engage in long-distance trade.[5] Some may perform the intellectual equivalent of raids on neighboring communities. Others will remain deeply ensconced in their original birthplaces. Still others may abandon the whole scene and join the circus. My point is not that these tribes should be dissolved; rather, it is that engagements of various kinds can encourage wide-ranging self-consciousness about the craft of political research. This, I believe, is for the better, whether one ultimately chooses to stick with longstanding tribal traditions or to forge new communities of inquiry.

Notes

[1] Thanks to Dvora Yanow, Timothy Pachirat, and Jan Kubik for their insightful comments on an earlier draft.

[2] In this highly stylized account, I put to the side the various other "tribes" one might consider, as well as a stunning variety of hybrid groups that have emerged across political science both in North America and beyond.

[3] This paragraph and the following one are based on work in progress (Schatz and Maltseva, n.d.) that involves coding the APSR for more than 100 years to see where (in the title, in the abstract, on the first page, and so on) each article first makes clear the country-context from which its evidence comes.

[4] Dvora Yanow (this symposium) argues that ethnographers may have certain preexisting personality traits that attract them to both the questions that they ask and the methods that they use. Thus, they may not intend to police boundaries.

[5] Thanks to Tim Pachirat for stimulating some of these metaphors.

References

Arias, Enrique D. 2009. "Ethnography and the Study of Latin American Politics: An Agenda for Research." In *Political Ethnography: What Immersion Contributes to the Study of Power*. Edward Schatz, ed. (Chicago: University of Chicago Press), 239–253.

Chen, Calvin. Forthcoming. "The Worm's Eye View: Using Ethnography to Illuminate Labor Politics and Institutional Change in Contemporary China." In *Knowing China: New Sources, Methods, and Field Strategies*. Mary Gallagher, Melanie Manion, and Allen Carlson, eds. (Cambridge: Cambridge University Press)

Geertz, Clifford. 1973. *The Interpretation of Cultures: Selected Essays*. New York: Basic Books.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor Books.

Ho, Karen. 2009. *Liquidated: An Ethnography of Wall Street*. Durham: Duke University Press.

Luke, Timothy. 1999. "The Discipline as Disciplinary Normalization: Networks of Research." *New Political Science* 21:3, 345–363.

Nagel, Thomas. 1989. *The View From Nowhere*. Oxford: Oxford University Press.

Scott, James C. 1985. *Weapons of the Weak: Everyday Forms of Peasant Resistance.* New Haven: Yale University Press.

Spiro, Melford E. 1990. "On the Strange and Familiar in Recent Anthropological Thought." In *Cultural Psychology: Essays on Comparative Human Development*. James W. Stigler, Richard A. Shweder, and Gilbert H. Herdt, eds. (New York: Cambridge University Press), 47–64.

Walsh, Katherine Cramer. 2004. *Talking about Politics: Informal Groups and Social Identity in American Life*. Chicago: University of Chicago Press.

Wedeen, Lisa. 1999. *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria*. Chicago: University of Chicago Press.

Announcements

Giovanni Sartori Award for the Best Book Published in 2008 Developing or Applying Qualitative Methods

Recipient: Margaret Somers, *Genealogies of Citizenship. Markets, Statelessness and the Right to Have Rights*. New York: Cambridge University Press, 2008.

Committee: Gary Herrigel, University of Chicago; Richard Ned Lebow, Dartmouth College; and Kathy Cramer Walsh, University of Wisconsin, Madison (chair).

The committee is very pleased to award Margaret Somers' *Genealogies of Citizenship. Markets, Statelessness and the Right to Have Rights* the 2009 Giovanni Sartori Award. The book analyzes the emergence (and perversion) of modern citizenship as a struggle between state, market, and civil society (public sphere) understandings of rights. The core argument is that over the last thirty years, particularly in the United States, the "right to have rights," which resides in civil society, has been continuously in retreat. On one side, states aggressively extend their scope of domination and autocratic control, while on the other side, market fundamentalism obsessively reduces human relations to contracts, exchanges, and structured sets of individual incentives. Both of these expansive realms have conceptions of rights that crowd out civil society-based conceptions based in non-contractual understandings of belonging, solidarity, and reciprocity—understandings that make possible the notion of a "right to have rights." This historical and conceptual argument is elaborately developed using a broad array of social and normative theories taken from across the social sciences. The theoretical argument is illustrated empirically in a series of rich and provocative cases, most notably the human rights tragedy that resulted in the aftermath of Hurricane Katrina in New Orleans.

In the committee's judgment, the book represents an exemplary application of qualitative methods in three ways.

First, it seamlessly weaves normative argument about what citizenship is and what it should entail, with empirical analysis of the way contending normative conceptions shaped the development of citizenship over the last several centuries. Further, the latter normative evidence is systematically deployed to bolster a sustained normative argument for reciprocal and non-contractual forms of sociality and citizenship.

Second, the book makes various kinds of narratives, myths, ideational regimes and cultural codes into the central empirical data of the work. These elements have an interesting status as "data" as they can be identified as practice, or in discourse, or as ideas about progress, or in symbols. Indeed, in using this data, Somers rejects the traditional coding of such things as only ideational or cognitive. She systematically shows that they have material dimensions in practice. The techniques used to identify these central empirical elements of her work are interpretive and empirical, observational, textual and archival. There is practical methodological creativity at every turn here. But the very notion that elements such as narrative and myth are data points is a methodological move for the construction of the argument: it organizes the way in which the empirical is identified and bolsters the argument about citizenship as the right to have rights.

Third, the work is an exemplary case of the genealogical method at work. As a method, genealogy systematically targets peripheral practices, ideologies, or discourses within dominant institutional and practical complexes. Somers uses the method to unearth narratives, myths, ideational frameworks, and cultural codes regarding citizenship as a form of belonging and reciprocal commitment in society that dominant state and market fundamentalist understandings and frameworks have marginalized. The claim is that such peripheral data create the possibility for the emergence of alternative constitutions, enactments and understandings of significant categories (such as citizenship).

In the committee's judgment, Somers' deft deployment of each of these three methods of qualitative analysis make the work into an exemplar for others seeking to do work of this kind in political science—and in the rest of the social sciences for that matter. It richly deserves the Sartori Award.

Alexander George Award for the Best Article or Book Chapter Published in 2008 Developing or Applying Qualitative Methods

Recipient: James Mahoney, Northwestern University, "Toward a Unified Theory of Causality." *Comparative Political Studies* 41:4–5 (April/May 2008), 412–436.

Committee: Mark Lichbach, University of Maryland; John Odell, University of Southern California (chair); and Benoît Rihoux, Université Catholique de Louvain, Belgium.

This profound article offers a way to translate causal language often used in case research into equivalent language in quantitative research and vice versa. Mahoney observes that analysts using case-oriented research often understand causality in terms of necessary and sufficient conditions, while analysts using population-oriented research understand causation as mean causal effect. The article concludes that the conception used in case-oriented research is appropriate for the population level, while the conception used in population research is useful for making predictions. Scholars in all fields of social science should read it.

SAGE Prize for the Best Paper Presented at the 2008 Annual Meeting of the American Political Science Association Developing or Applying Qualitative Methods

Recipient: Robert Adcock, George Washington University: "The Curious Career of the 'Comparative Method': The Case of Mill's Methods."

Committee: Carsten Q. Schneider (chair), Central European University, Hungary; Michael Bernhard, University of Florida, Gainesville; Edward Schatz, University of Toronto, Canada.

Robert Adcock provides an innovative account of John Stuart Mill's comparative method, a logic of inquiry that is often used and criticized in present day comparative social science literature. Based on an exceptionally thorough reading of Mill's work—well beyond the sections in which his famous methods of agreement (MoA) and method of difference (MoD) are outlined—Adcock skillfully demonstrates that both adherents and adversaries of Mill often get it wrong: the former, because they oversimplify and often just refer to these two methods as some sort of methodological fig-leaf, the latter because they fail to recognize that Mill himself not only formulated caveats about MoA and MoD that sound very similar to present day critiques, but he also formulated a set of alternative research approaches more suitable for the study of social and political phenomena. Many scholars, including the award committee, have always been a little wary of the fact that we tend to ignore the method of concomitant variation, that we seemed to ignore Mill's warnings about the limitation of his methods for the study of society, and that certain researchers used this as reason to be skeptical of small-n cross national research per se. After reading the Adcock piece, these oversimplifications of Mill are no longer easily justifiable. If widely read, new generations of social scientists will come to realize that 'Mill's methods' are more than MoA and MoD and include the Method of Concomitant Variation and the Method of Residues, a fairly unknown part of Mill's system of logic about which Adcock promises to write more in the near future.

The committee also deems it a methodological advancement that Adcock's paper forcefully demonstrates how fruitful it can be to go back and read the classics. We learn that Mill had interesting things to say that are (still) en vogue in present day methodological debates, such as the vices and virtues of increasing the number of cases, omitted variable bias, case selection principles, or the meaning and consequences of experimental ('directly inductive') versus 'deductive' sciences.

The committee also suggests bestowing an "honorable mention" to Leah Gilbert and Payam Mohseni for their paper on 'Contested Concepts: Mapping the Boundaries of Hybrid Regimes.' The authors are still graduate students, which makes their achievement even more remarkable: providing a carefully argued suggestion on how to think about and conceptualize those political regimes that are neither democratic nor authoritarian and to do so in the form of a paper that is highly convincing, useful, and pleasant to read.

Qualitative and Multi-Method Research
School of Government and Public Policy
University of Arizona
315 Social Sciences Building
P.O. Box 210027
Tucson, AZ 85721-0027
USA

Qualitative and Multi-Method Research is edited by Gary Goertz (tel: 520-621-1346, fax: 520-621-5051, email: ggoertz@u.arizona.edu). The assistant editor is Joshua C. Yesnowitz (email: jcyesnow@bu.edu). Published with financial assistance from the Consortium for Qualitative Research Methods (CQRM). Opinions do not represent the official position of CQRM. After a one-year lag, past issues will be available to the general public online, free of charge, at http://www.maxwell.syr.edu/moynihan/programs/cqrm/section.html. Annual section dues are $8.00. You may join the section online (http://www.apsanet.org) or by phone (202-483-2512). Changes of address take place automatically when members change their addresses with APSA. Please do not send change-of-address information to the newsletter.