

Qualitative & Multi-Method Research

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

Contents

Symposium: Case Selection, Case Studies, and Causal Inference

<i>Introduction</i>	
David Collier	2
<i>Do the N's Justify the Means?</i>	
David A. Freedman	4
<i>Response to David Freedman</i>	
James D. Fearon and David D. Laitin	6
<i>Techniques for Case Selection: A Response to Freedman</i>	
John Gerring	7
<i>Choosing Cases for Case Studies: A Qualitative Logic</i>	
Gary Goertz	11
<i>Rejoinder</i>	
David A. Freedman	14

Symposium: Field Experiments and Qualitative Methods

<i>Natural and Field Experiments: The Role of Qualitative Methods</i>	
Thad Dunning	17
<i>The Promising Integration of Qualitative Methods and Field Experiments</i>	
Elizabeth Levy Paluck	23
<i>Battling Onward: The Debate Over Field Research in Developmental Economics and its Implications for Comparative Politics</i>	
Edmund J. Malesky	30
<i>The Contribution of Area Studies</i>	
Stephen E. Hanson	35
Book/Article Notes	43
Announcements	45

APSA-QMMR Section Officers

President: John Gerring, Boston University
 President-Elect: Colin Elman, Syracuse University
 Vice President: Margaret Keck, Johns Hopkins University
 Secretary-Treasurer: Colin Elman, Syracuse University
 Newsletter Editor: Gary Goertz, University of Arizona
 Division Chair: Rudra Sil, University of Pennsylvania
 Executive Committee: Peri Schwartz-Shea, Univ. of Utah
 Rose McDermott, University of California, Santa Barbara
 Page Fortna, Columbia University
 Dan Carpenter, Harvard University

Letter from the Editor

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

The Qualitative and Multi-Methods Section had another good showing at the APSA convention in Boston. Thanks from the section go to Craig Parsons and Hillel Soifer for organizing an excellent set of panels. Informal impressions from the organizers and others (including myself) were that panels were well attended, even though in some cases we had three panels in the same time slot. Paper proposals were definitely up this year. This is important for the section since the number of proposals is an important part of the formula used by APSA to allocate panels, and is hopefully also a sign of increased research in qualitative methods. Rudy Sil (rudysil@sas.upenn.edu) will be organizing panels for APSA 2009; contact him with your ideas.

Increased research needs publication outlets. This is particularly critical for graduate students and untenured assistant professors. As such it is very good that Jim Caporaso as editor of *Comparative Political Studies* has reported (see the Announcements section of the newsletter for more) the creation of a "Methodology Forum" as a part of the journal. While *CPS* has become a central journal for qualitative methods it is still nice to see official recognition of this on the part of the journal. I also note that *Political Research Quarterly* has become an important outlet for qualitative methods articles. The Announcements section also gives details on a new methods series (Palgrave Macmillan) which welcomes and encourages qualitative methods submissions.

A new APSA Conference Group on Interpretive Methodologies & Methods has been formed to provide a forum for the discussion of methodologies and methods related to interpretive research. The Announcements section provides details and a URL. This is just another sign of the expanding interest in different methodologies. The next issue of the newsletter will have a symposium on teaching interpretive methodologies, which I think will be extremely useful.

The Institute for Qualitative and Multi-Method Research is making a major move both in location and dates. Since its inception it has been held in January at Arizona State University. We have exchanged the sunny winter skies of Phoenix for the reportedly fine weather of late-spring upstate New York. The Institute will now be held in late May and early June and

will be hosted by the Maxwell School, Syracuse University. See the Announcements section for details, new website, deadlines for application, etc.

The *Oxford Handbook of Political Methodology* was published just in time for APSA 2008. Janet Box-Steffensmeier, Henry Brady, and David Collier have done an incredible job of putting together 37 chapters on methodological issues. There are at least 15 chapters that are directly relevant to qualitative methods (see the Announcements section for a list). I am teaching a graduate qualitative methods seminar this semester and have found this handbook a very useful source of readings.

It is probably not surprising that chapters in the *Oxford Handbook* have already attracted controversy. The “Case Selection, Case Studies, and Causal Inference” symposium in this newsletter leads off with David Freedman’s critique of the Fearon-Laitin and Gerring chapters of the *Handbook*. Issues surrounding case study methodology continue to provoke much discussion within the section, and have been a major topic in newsletters. This symposium addresses the core issue of selecting cases for intensive case analysis.

There has been a trend over the last 15–20 years to devalue “area studies” and along with that field research and

country-specific knowledge. In another contribution to this issue of the newsletter, Steve Hanson addresses many of the critiques of “area studies”—along with field research and country-specific knowledge. The symposium “Field Experiments and Qualitative Methods” argues that many of the core hypotheses of political science can be tested by country experts using designed (by the researcher) or natural (researcher does not control) experiments. Dunning’s contribution gives a nice example of how this works in one case in exploring classic hypotheses about crosscutting cleavages with an experiment in Mali. Paluck discusses how qualitative analysis, field experiments, and area knowledge could be profitably integrated. Malesky discusses how field experiments in developmental economics have challenged traditional large- N , cross-national, regression-type analyses. He also gives a nice summary of the strengths and weaknesses of this kind of research. One thing that this symposium implies is that graduate students need to be on the look-out for natural experiments as they go into the field; they often occur in unexpected places and ways. If the researcher has her eyes open, she can take advantage of these opportunities when they arise.

Symposium: Case Selection, Case Studies, and Causal Inference

David Freedman, author of the opening and closing contributions to this symposium, passed away on October 17, 2008. A Professor of Statistics at the University of California, Berkeley, Freedman strongly believed that case knowledge and qualitative evidence are crucial to causal inference. An important statement of his view, noted elsewhere in this issue of the newsletter, is found in Freedman, “On Types of Scientific Enquiry: The Role of Qualitative Reasoning” (*Oxford Handbook of Political Methodology*, 2008).

Introduction

David Collier

University of California, Berkeley
dcollier@berkeley.edu

For scholars concerned with causal inference, how should cases be selected in case study research?

This symposium builds on previously published arguments by James Fearon and David Laitin (2008), who favor random sampling in case study analysis, and by John Gerring (2008), who favors purposive selection. The statistician David Freedman—long an advocate of case studies as an important research tool—comments on these published arguments; responses are offered by Fearon-Laitin and by Gerring; Gary Goertz adds a commentary of his own; and then Freedman offers concluding remarks.

In Fearon and Laitin’s (2008) discussion, the goal is to draw insights about causal mechanisms from case studies so as to illuminate the findings from a large- N , regression-type analysis. The idea of random sampling is of course central to the broad literature on statistical inference, and for Fearon and Laitin a key advantage of this approach is to prevent scholars from deliberately selecting cases favorable to their preferred hypotheses, thus engaging in “cherry-picking.”

By contrast, in advocating purposive selection Gerring (2008) draws on the tradition that reaches back at least to understandings of case studies offered by Lijphart (1971), Eckstein (1975), and George (1979). Gerring’s approach employs a large- N framework, which he uses to identify cases that are seen as typical, diverse, extreme, deviant, influential, crucial, pathway, most similar, and most different.

Yet another perspective, introduced in this symposium by Gary Goertz, likewise advocates purposive selection for case-study research aimed at causal inference. Goertz is primarily interested in the case studies in their own right, rather than their role in statistical analysis involving a large N . Goertz’s point of departure is the cross-tabulation of two dichotomies (the outcome to be explained and the potential explanation), and his discussion of case selection focuses on choices among the cells in the resulting 2 x 2 table. This approach connects with the wider tradition of analyzing matching and contrasting cases, identified in different ways with the methods of agreement and difference of J. S. Mill (1974 [1843]), most-similar and most-different designs of Przeworski and Teune (1970), and Qualitative Comparative Analysis (Ragin 1987; see also 2000).

Freedman extends, refines, and in some ways departs from the above approaches. His overall position is to prefer purposive selection. For case-study analysis concerned with check-

ing models employed in large-*N* research, he recommends a focus on cases consistent with predictions of the model, cases not consistent with its predictions, and influential cases that appear to have an especially strong effect on findings derived from the model.¹

Among the issues discussed in this symposium, I find three to be of special interest. First, the idea of random sampling from a well-defined population is a gold standard for descriptive inference, and quite properly so. However Freedman suggests that this standard is less frequently—and less effectively—met than is often believed. Observational studies in the social sciences often involve some variant of a convenience sample. This certainly would appear true in macro-comparative research, as with a focus on the OECD countries. In that instance, one may have a convenience sample (driven in part by the availability of excellent data), but the presumed population of interest may never be clearly defined. Even with a random sample, Freedman points out that major problems of missing data can weaken inferences from sample to population. A statistical model may be necessary in correcting for potential bias due to missing data, yet this model may well introduce more bias than it removes. Freedman argues that as a consequence, a random sample can pose just as many uncertainties about generalization as a convenience sample.

In my view, the better part of wisdom may be to recognize that, under some (possibly many) circumstances, we should drop the pretense that we engage in random sampling from a defined population. Being realistic, departures from this (obviously useful) gold standard occur frequently. These considerations should, at least some of the time, lead scholars to be more cautious about undertaking generalization, and the expression “external validity” may sometimes raise higher expectations for achieving valid generalization than are warranted or appropriate. It is often more productive to pursue contingent generalizations by seeking to map findings from a particular set of cases onto carefully specified additional cases (possibly including, in international studies, additional world regions).

Second, Freedman agrees that cherry-picking should be avoided. However, he notes that until the scholar has actually done the case study research, it is often hard to know how cases will come out. This uncertainty makes it less likely that the researcher can intentionally select cases that support a preferred hypothesis. I am reminded of Donald Campbell’s (1975) argument that the findings of case studies routinely go in a different direction than the researcher expects before starting the investigation. Cherry-picking may thus not be as grave a problem as the vivid metaphor suggests.

These comments about our weak prior knowledge of how particular cases will actually come out are certainly relevant to Goertz’s focus on selecting cases from particular cells within his 2 x 2 table. How does one know in which cell the cases will be located? One solution is suggested by Gerring’s approach. He uses large-*N* regression-type analysis—based on what is doubtless a more preliminary and imprecise coding of cases—in initially situating the cases; this is subsequently to be followed by the fine-grained coding that researchers can achieve,

based on their close case knowledge.

Third, Freedman has long argued that descriptive findings are too often interpreted as causal relationships, with far too little attention to the fragility of causal inference. Correspondingly, descriptive findings may be given an importance that, taken by themselves, they do not deserve.

This perspective leads Freedman to note with concern the opening statement in Fearon-Laitin (2008), where they say that “almost by definition, a single case study is a poor method for establishing whether or what empirical regularities exist across cases. To ascertain whether some interesting pattern, or relationship between variables, obtains, the best approach is normally to identify the largest feasible sample of cases relevant to the hypothesis....”

Freedman comments that if these empirical regularities and relationships among variables are of interest because they contribute to causal inference, then this approach is too rigid, inappropriately devalues case studies, and fails to recognize the very different paths that can be followed in inferring causation. Freedman sees case studies as making diverse contributions: they can “overturn prior hypotheses, generate new lines of inquiry, or confirm causal claims.” The empirical regularities that emerge in case studies may lack the presumed generality of those derived from large-*N* analysis. Yet they may contribute just as much because they rest on what may be the considerably greater power of insight derived from close case knowledge.

In conclusion, as Freedman puts it in his final remarks, this debate “has a happy ending.” Any apparent disagreement with Fearon and Laitin over random selection and case studies is resolved through the exchange in this symposium. More broadly, there would doubtless be a consensus among the contributors that, as Freedman puts it,

- (i) There are many ways to do good science. (ii) In particular, neither cluster of methods has a general advantage over the other. (iii) Therefore, there are many fruitful ways for qualitative and quantitative researchers to interact.

Standards can and should be applied in evaluating alternative causal claims, but there is certainly no single method through which this analytic task should be accomplished.

Notes

¹ Here and elsewhere in this symposium, “model” refers to a statistical model, and should not be confused with a game-theoretic model. A statistical model is understood as a set of one or more mathematical equations—commonly regression equations—used in the analysis of empirical data. Among many purposes, a model may be employed in descriptive inference, as in an inference from a sample to a population, and in causal inference. As Freedman emphasizes in this symposium, descriptive inference faces numerous challenges, and he has argued in many publications that causal inference based on statistical models is very fragile indeed. Both of these points are crucial to the present discussion.

References

- Box-Steffensmeier, Janet M., Henry E. Brady, and David Collier, eds. 2008. *Oxford Handbook of Political Methodology*. Oxford: Oxford University Press.
- Campbell, Donald T. 1975. "'Degrees of Freedom' and the Case Study." *Comparative Political Studies* 8:2 (July), 178–93.
- Eckstein, Harry. 1975. "Case Study and Theory in Political Science." In Fred I. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science*, Vol. 7. Reading, MA: Addison-Wesley.
- Fearon, James D. and David D. Laitin. 2008. "Integrating Qualitative and Quantitative Methods." In *The Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (New York: Oxford University Press), 756–76.
- Freedman, David A. 2008. "On Types of Scientific Enquiry: The Role of Qualitative Reasoning." In *The Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (New York: Oxford University Press), 300–18.
- George, Alexander L. 1979. "Case Studies and Theory Development: The Method of Structured, Focused Comparison." In Paul Gordon Lauren, ed., *Diplomacy: New Approaches in History, Theory and Policy*. New York: Free Press.
- Gerring, John. 2008. "Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques." In *The Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (New York: Oxford University Press), 645–84.
- Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65:3 (September): 682–93.
- Mill, John Stuart. 1974b [1843]. "Of the Four Methods of Experimental Inquiry." In Book 3, Chapter 8, *A System of Logic, Ratiocinative and Inductive*. Toronto: University of Toronto Press.
- Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: John Wiley.
- Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of

Do the *N*'s Justify the Means?

David A. Freedman
University of California, Berkeley

Box-Steffensmeier, Brady, and Collier (2008) examine the craft of political science from a rich variety of perspectives. I will comment on two chapters, one by Fearon and Laitin, the other by Gerring. These chapters are well reasoned, but reach opposite conclusions about a basic issue—how should cases be chosen? Fearon and Laitin focus on large-*N* research, with a logit model for civil war to illustrate the argument. To see whether causal inferences from the model hold up under closer scrutiny, they choose a sample of cases for detailed investigation ("multi-method research").

Fearon and Laitin say, "An important but neglected problem for this research approach is the question of how to choose the cases for deeper investigation.... We propose that choosing cases for closer study *at random* is a compelling complement in multi-method research to large-*N* statistical methods in its ability to assess regularities and specify causal mechanisms" (758).

By contrast, according to Gerring, "In order to isolate a sample of cases that both reproduces the relevant causal features of a larger universe (representativeness) and provides variation along the dimensions of theoretical interest (causal leverage), case selection for very small samples must employ purposive (non-random) selection procedures" (645).

In short, Fearon and Laitin recommend sampling cases at random, whereas Gerring recommends purposive selection. To be sure, Gerring's main interest is choosing cases for small-*N* research, but his reasoning applies equally well to the multi-method research discussed by Fearon and Laitin. I will not resolve the conflict here, although I will make some suggestions. The essays raise other important questions about research methodology, and I will also comment on those.

At the outset, Fearon and Laitin make three valuable points. (1) Scholars can be remarkably, let's say, innocent when describing research designs and case selection. (2) "Cherry-picking cases" (by which Fearon and Laitin mean picking cases that support a particular line of argument) is often a bad idea. (3) Random sampling precludes cherry-picking.

An emphasis on choosing cases purely at random, however, may be misplaced. By now, fitting models to data is routine, and there are any number of well-intentioned software packages that automate large parts of the activity. With a totally random sample of cases, the likely finding is that the sample follows the trends predicted by the model.¹ After all, large-*N* scholars choose models that do a good job of tracking the data. (And if the first model they try doesn't work, they might go for a second model—or a third.)

When Fearon and Laitin get down to business, they choose a stratified random sample—stratified not only by explanatory variables (region) but also by the outcome variable (presence or absence of civil war). So the methodological advice amounts to this: within strata, choose your cases at random.

The advice is excellent, if you have a lot of cases in each stratum, and can afford a sample of reasonable size. But how does it help someone who does qualitative research where the number of cases is strictly limited? On the other hand, for model-checking in large-*N* research, I would recommend taking (i) some cases that are consistent with the predictions of the model, (ii) some that are inconsistent, and (iii) some from strata of special interest. Cases that markedly influence results should be considered too. Finally, random sampling is good and cherry-picking is bad—unless, of course, you want to make an existence proof or an argument *a fortiori*.

Fearon and Laitin conclude that studying the sample cases is a useful extension of the statistical modeling and suggests "a natural way that qualitative work might be integrated into a research program as a complement to rather than as a rival or substitute for quantitative analysis" (774–75). Folding qualitative work into a quantitative research program is an idea, but a general recommendation seems premature—especially when the evidence consists of a case study with *N* = 1, namely, their own investigation of civil war.

Fearon and Laitin also make an interesting comparison between small-*N* and large-*N* research methods: "Almost by

definition, a single case study is a poor method for establishing whether or what empirical regularities exist across cases. To ascertain whether some interesting pattern, or relationship between variables, obtains, the best approach is normally to identify the largest feasible sample of cases relevant to the hypothesis or research question, then to code cases on the variables of interest, and then to assess whether and what sort of patterns or associations appear in the data” (757, footnote omitted).

The claimed superiority of large-*N* methods is obviously right if “empirical regularities” are statistical measures of association, like regression coefficients. The thesis is less obvious if “empirical regularities” are defined more broadly, so as to include (for example) causal relationships. Then Fearon and Laitin’s “best approach” seems too rigid. In fact, a lot of good science gets done rather differently (Freedman 2008a).

Therefore, I suggest taking a more liberal view of the relationship between qualitative and quantitative research. Qualitative methods can overturn prior hypotheses, generate new lines of inquiry, or confirm causal claims. Indeed, large-*N* research is often done to confirm insights generated by case studies (Freedman 2008a). It should be common ground, however, that the best research programs combine qualitative and quantitative methods.

Looking beneath the surface of a statistical model is hard work, and requires intellectual fortitude for that reason among others. Fearon and Laitin looked, using an elegant and systematic technique, and reported what they saw. This might be an example worth following.

I turn now to Gerring. His Table 1 lists a variety of methods for case selection. Most of the suggestions are helpful, as is the accompanying discussion. However, some entries in the table are puzzling. For example, the table recommends the hat matrix and Cook’s distance, which measure in different ways how each observation influences the regression outputs. Such measures might be helpful when selecting cases to probe a large-*N* model. For qualitative research, however, regression output seems irrelevant. The table also recommends discriminant analysis and factor analysis. But these are large-*N* techniques, pure and simple—or impure and madly complicated, depending on one’s perspective.²

Gerring proceeds to make a strong claim about case selection: “The most useful statistical tool for identifying cases for in-depth analysis in a most-similar setting is probably some variety of matching strategy—e.g., exact matching, approximate matching, or propensity-score matching” (670, footnote omitted).

It is hard to see how techniques like propensity-score matching apply to small-*N* research.³ Even for large-*N* research, the claim ignores abundant evidence on the fallibility of matching techniques. I agree that matching may have a role to play, but suggest that caution is in order.

Gerring raises broader issues that should be addressed too. For example, he says: “In large-sample research, the task of case selection is usually handled by some version of randomization” (645). I disagree. *Some* large-*N* research is based on randomized experiments or probability samples, but most is

not. Convenience samples and observational studies are far more typical, with statistical models to address selection effects and confounding.

The difficulties with the modeling approach are well known (Berk 2004; Brady and Collier 2004; Freedman 2005; Mahoney and Rueschemeyer 2003). Of course, there will always be those who can ignore the difficulties. See, for instance, King, Keohane, and Verba (1994).

Gerring also has something noteworthy to say about experiments: “[i]n a randomized experiment... the researcher typically does not attempt to measure all the factors that might affect the causal relationship of interest. [ii] She assumes, rather, that these unknown factors have been neutralized across the treatment and control groups by randomization or by the choice of a sample that is internally homogeneous” (670).

Point (i) is correct. Point (ii) is off the mark. If the experiment is properly done, few assumptions are needed, because randomization guarantees that the treatment and control groups are balanced on the average. That is why experiments give unbiased estimates of causal effects.⁴ Furthermore, there is no need to choose “a sample that is internally homogeneous”—which is all to the good, since that task is beyond our present capabilities.

When the computer actually prints out the random numbers that define the treatment and control groups, there will be minor imbalances due to the play of random chance. These imbalances are the source of random errors in estimates derived from the data. The impact of random errors is conventionally measured by standard errors and *P*-values. It is the randomization that justifies the conventional measures. Without the randomization, justification might be elusive.⁵

The *intention-to-treat principle* is to compare rates or averages for those assigned to treatment with those assigned to control. That is the tacit premise of the discussion. If regression adjustments are made to compensate for imbalances between groups, or to correct for crossover, matters become substantially more complicated (Freedman 2006, 2008b, 2008c, 2008d).

The logic of randomized controlled experiments is worth understanding, for two reasons at least. (i) Experiments are the gold standard for causal inference. (ii) The statistical methods used to analyze observational studies usually depend on the assumption that in some respect or another, the observational study at hand is like an experiment. The logic of randomized controlled experiments is therefore central, even for observational research.

How to choose cases? This question has intrigued scholars from John Stuart Mill onwards, perhaps because the answer depends on context. Any additional clarity is to be welcomed, and Gerring has provided more than a little. So have Fearon and Laitin.

On the other hand, some of the methods that Gerring proposes are ill-suited to qualitative research. Furthermore, he mistakes the role of random sampling and experimentation in large-*N* research, and fails to recognize the limits of other large-*N* techniques. Fearon and Laitin seem at times to imply that qualitative methods are useful only as checks on quantitative

results. Such a perspective would undervalue contributions made by small-*N* methods. More generally, that kind of perspective ignores a crucial point: there are many ways to do good science.

Notes

¹ Fearon and Laitin show that close examination of typical cases (countries with no civil war and low probability of civil war according to the model) can be illuminating—a special and valuable feature of their research. Indeed, they use the cases to check the qualitative implications of their causal model.

² Seawright and Gerring (2008) give a clearer account of the matter, indicating that the relevant population must be large.

³ The setting for propensity-score matching is usually an observational study where subjects self-select into one of two conditions; call these “treatment” and “control.” The first step is usually to estimate the conditional probability that a subject winds up in treatment, given the covariates. Logit models are often used. This is not an activity to be undertaken with a small sample. For empirical evidence on the weaknesses of matching designs in large-*N* research, see for instance Arceneaux, Gerber, and Green (2006), Glazerman, Levy, and Myers (2003), Peikes, Moreno, and Orzol (2008), Wilde and Hollister (2007). For additional discussion pro and con, see *Review of Economics and Statistics* 86:1 (February 2004); *Journal of Econometrics* 125:1–2 (March–April 2005).

⁴ Suppose, for instance, that we have an experimental population of 1,000 subjects, with 400 chosen at random and assigned to treatment; the remaining 600 are the controls. Each subject has two potential responses: one is observed if the subject is assigned to treatment, and the other if assigned to control. The average response of the 400 is an unbiased estimate of what the average would be if all 1,000 subjects were assigned to treatment. Likewise, the average response of the 600 is an unbiased estimate of the average response if all 1,000 subjects were controls. The general principle is this: with a simple random sample, the sample average is an unbiased estimate of the population average. For additional details, see Freedman (2006).

⁵ For example, see Freedman, Pisani, and Purves (2007). Chapter 27 discusses experimental comparisons; technical detail is provided in A31–36. Chapter 29 explains what happens without randomization; also see Freedman (2008e).

References

- Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2006. “Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment.” *Political Analysis* 14: 37–62.
- Berk, Richard A. 2004. *Regression Analysis: A Constructive Critique*. Sage Publications.
- Box-Steffensmeier, Janet M., Henry E. Brady, and David Collier. 2008. *The Oxford Handbook of Political Methodology*. New York: Oxford University Press.
- Brady, Henry E. and David Collier. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman and Littlefield.
- Fearon, James D. and David D. Laitin. 2008. “Integrating Qualitative and Quantitative Methods.” In *The Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (New York: Oxford University Press), 756–76.
- Freedman, David A. 2005. *Statistical Models: Theory and Practice*. New York: Cambridge University Press.
- Freedman, David A. 2006. “Statistical Models for Causation: What Inferential Leverage Do They Provide?” *Evaluation Review* 30:

691–713.

- Freedman, David A. 2008a. “On Types of Scientific Enquiry: The Role of Qualitative Reasoning.” In *The Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (New York: Oxford University Press), 300–18.
- Freedman, David A. 2008b. “On Regression Adjustments in Experiments with Several Treatments.” *Annals of Applied Statistics* 2: 176–96.
- Freedman, David A. 2008c. “Randomization Does Not Justify Logistic Regression.” *Statistical Science* 23: 237–49.
- Freedman, David A. 2008d. “Survival Analysis: A Primer.” *The American Statistician* 62: 110–19.
- Freedman, David A. 2008e. “Oasis or Mirage?” *Chance* 21: 59–61.
- Freedman, David A., Robert Pisani, and Roger A. Purves. 2007. *Statistics*. 4th edition. New York: W. W. Norton & Company, Inc.
- Gerring, John. 2008. “Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques.” In *The Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (New York: Oxford University Press), 645–84.
- Glazerman, Steven, Dan M. Levy, and David Myers. 2003. “Non-experimental versus Experimental Estimates of Earnings Impacts.” *Annals of the American Academy of Political and Social Science* 589: 63–93.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Mahoney, James and Dietrich Rueschemeyer. 2003. *Comparative Historical Analysis in the Social Sciences*. New York: Cambridge University Press.
- Peikes, Deborah N., Lorenzo Moreno, and Sean Michael Orzol. 2008. “Propensity Score Matching: A Note of Caution for Evaluators of Social Programs.” *The American Statistician* 62: 222–31.
- Seawright, Jason and John Gerring. 2008. “Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options.” *Political Research Quarterly* 61: 294–308.
- Wilde, Elizabeth T. and Robinson Hollister. 2007. “How Close Is Close Enough? Evaluating Propensity Score Matching Using Data from a Class Size Reduction Experiment.” *Journal of Policy Analysis and Management* 26: 455–77.

Response to David Freedman

James D. Fearon
Stanford University
jffearon@stanford.edu

David D. Laitin
Stanford University
dlaitin@stanford.edu

As Jonathan Swift made mockingly clear, “modest proposals” that purport to solve previously unyielding problems can have horrible implications. Such proposals should be subjected to skeptical analysis. So we are pleased that our proposed random method of case selection for the qualitative component of multi-method research has attracted some skeptical commentary in the research community in political science.¹ And we are very grateful to David Freedman for providing a perspective on our approach. He is especially qualified to do

so, as he is a leading statistician who has long worried about inflated claims for statistical methods in the social sciences, and has been a champion of approaches that are sensitive to the particularities of each datapoint.

We completely agree with Freedman's claim that there are many ways to do good social science. Indeed, as Freedman quotes us, we argued that the random narratives approach is "a compelling complement" to large-*N* research. This is more modest than Freedman's implication that we believe we have discovered the one true path for multi-method research. In fact, if everyone did random narratives, there would be no expert narratives for the research community to consult!

Furthermore, as Freedman points out, the method has been applied only to our work on civil war onsets. Perhaps it will not be the best approach for other questions that scholars want to use multiple methods to address. We agree, although one goal of our article was to argue that there are good *a priori* (or theoretical) reasons to think that the approach could be valuable for research designs on topics other than civil war onset. Multi-method and other social science research inevitably involves a process of going back and forth between theory and data (despite the pristine hypothesis-testing scenario assumed in statistics textbooks). The random narratives approach is a way to discipline and make more productive this back-and-forth process in a fairly typical social science setting, where one has cross-sectional or panel data with which to document empirical patterns, and historical materials available to investigate causal mechanisms in particular cases.

In the case of our work on civil war, we constructed a country/year dataset with civil war onset as the dependent variable.² We estimated a statistical model that identified several correlates of civil war onsets for which we proposed possible causal interpretations. The interpretations were based on a reading of the statistical results *and* our previous knowledge of a set of cases well known to us. To look at those same cases for qualitative support for a causal interpretation would have been intellectual double-dipping. The method of randomly selecting cases for analysis of causal mechanisms behind peace or war onset helped us to avoid or at least reduce this bias. But we certainly do not maintain that random selection of cases for detailed analysis would always be the most effective and efficient approach in a multi-method research project, independent of the subject matter or the stage of the research (in terms of "back and forth").

Freedman writes that our "claimed superiority of large-*N* methods is obviously right if 'empirical regularities' are statistical measures of association, like regression coefficients. The thesis is less obvious if 'empirical regularities' are defined more broadly, so as to include (for example) causal relationships." We agree here as well, although we were trying in the cited sentences precisely to distinguish empirical regularities in the sense of mere associations from causal relationships. We would not claim a generalized superiority of large-*N* methods for identifying causal relationships. Indeed, the main idea of the multi-method approach we are endorsing is to use case-specific evidence systematically to assess whether causal in-

terpretations of the mere associations seen in a regression analysis are justified.

We do not therefore see how Freedman attributes to us the notions of the "superiority" of large-*N* methods or that "qualitative methods are useful only as checks on quantitative results." These claims may suggest incorrectly that we think causal relationships are easily read out of large-*N* statistical studies in social science. They also misread our view of the contributions made by small-*N* methods in the overall research process. In practice, as we noted above, there is a constant back-and-forth between data and theory in social science research, with case study evidence entering in more than one way. Knowledge of particular cases often helps to suggest causal mechanisms that may or may not be common and relevant in a larger sample of cases, and so may motivate and guide construction of a large-*N* study. A large-*N* study may in turn reveal new and different-from-expected patterns that stimulate new (or revised) theorizing about causal relationships, which may then be assessed by a return to case studies (chosen at random?). Those case studies may suggest new causal relationships that can subsequently be put to test with a newly constructed dataset. So it often happens in political science that case studies come into the scientific process at an early stage, motivating the research and the source of early conjectures, and then again at a later stage, after the regressions have been run.

Researchers in comparative politics invariably go back-and-forth between theory and data, and quite often they go back and forth between cases and broad patterns. Our modest proposal is an attempt to make progress on the question of by what principles to choose the cases in the context of the back and forth.

Notes

¹ See for example Evan Lieberman's critique of our proposed method, "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99:3 (August 2005), 435–52.

² James D. Fearon and David D. Laitin, "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97:1 (February 2003), 75–90.

Techniques for Case Selection: A Response to David Freedman

John Gerring
Boston University
jgerring@bu.edu

Recognition of the problem posed by case selection in case study research stretches back, arguably, to the very beginnings of the genre, e.g., to early work by Frederic Le Play (1806–1882) and Florian Znaniecki (1882–1958). Harry Eckstein's (1975) classic study, a point of departure for political scientists today, appeared over three decades ago. Clearly, the field has been struggling with this issue for some time.¹

The objective of my chapter for the *Oxford Handbook* was to summarize extant approaches to case selection, to add some new techniques to this battery of strategies (or at least provide a moniker for and a formal treatment of techniques that are already widely practiced), and to show how quantitative techniques might be brought to bear on these matters. (“Statistical” and “quantitative” will be employed synonymously in this discussion.) With respect to the latter, my argument was that most case-selection techniques could be practiced either qualitatively or quantitatively, given the right circumstances. (One technique, the “crucial case,” can be practiced only qualitatively.) Nine purposive case-selection procedures were reviewed, each associated with a distinct case-study type: *typical*, *diverse*, *extreme*, *deviant*, *influential*, *crucial*, *pathway*, *most-similar*, and *most-different*.

I am very grateful to David Freedman—whose work has served as a touchstone for many of us—for offering his thoughts on this long-running issue and for offering me an opportunity to clarify and refine my initial statement. I believe that a good deal of common ground can be located here.

I should say, at the outset, that my chapter was based on prior work which presented these issues at greater length (Gerring 2007a, 2007b; Seawright and Gerring 2008). The chapter in the *Oxford Handbook* was not intended to provide a comprehensive treatment of this very large subject. Indeed, some considerations important to case selection were omitted entirely from the chapter, for lack of space (e.g., Gerring and McDermott 2007). Quite possibly, some of the points of apparent disagreement are a product of the condensed format demanded by the *Handbook*. Others are doubtless due to my own failure of communication. There may also be one or two points of genuine disagreement. In any case, I am anxious to explore what these might be, in the hopes that by doing so we can move the field forward.

Let me begin with what I take to be a point of agreement. Large-sample work aims for some version of random sampling. This is the textbook method of case selection. Freedman corrects my overly optimistic assessment, pointing out that large- N research often does not achieve this aim. Although I have not studied the matter, I would assume that random sampling is often achieved when the units of analysis are individuals and when these responses are drawn from survey research (e.g., in public opinion studies). I would assume that random samples are generally not achieved when the individuals composing a sample are being subjected to an experimental protocol or when the units of theoretical interest are larger entities such as countries.

By contrast, research based on very small samples ($N=1$ or several) cannot employ this time-honored technique of case selection for two critical reasons. First, there is a high likelihood that the (randomly) chosen cases will be wildly unrepresentative of the population (note that where $N=5$, sampling variance is much higher than where $N=50$ or 100). Second, a randomly chosen sample is unlikely to provide adequate leverage for the research question under investigation (note that where N is large, the resulting sample is likely to contain variation on theoretically relevant dimensions, but this is not

the case where N is very small). This was not an issue addressed in my chapter for the *Handbook*, but it is an important assumption underlying the chapter (Gerring 2007a; Seawright and Gerring 2008), and one that may be worth expatiating upon in the present context.

To clarify, where chosen samples are medium to large, as in Fearon and Laitin’s sample of 25 countries, it is reasonable to employ random sampling or stratified random sampling, as they do. (Note, however, that because one of Fearon/Laitin’s sampling criteria is the existence of a civil war—the dependent variable of interest—it does not fit the usual understanding of stratified random sampling. Still, there are often good reasons for selecting on the dependent variable, in the tradition of case-control studies.)

Naturally, the existence of 25 country-cases imposes a considerable burden on the analysts, necessitating long, in-depth studies for each case (which the authors are in the process of completing). Typically, when the number of chosen cases is this extensive, the amount of detail and original research devoted to each case is limited. Depth and breadth tend to vary inversely. This recalls a point of definition. If a “case study” refers to an in-depth analysis of a single case (with the objective of saying something about a broader population of cases), then the case study format becomes more diffuse as N increases. However, there is no hard and fast boundary between a case study and a cross-case study. One flows into the other. That is one of the many ambiguities of the term (see below).

Now, let us suppose counterfactually that Fearon and Laitin proposed to conduct a study of a single case (e.g., Algeria), or a very small sample composed of three or four cases, while maintaining their theoretical interest in generating insights about a global population of nation-states. Here, there are lots of reasons to be suspicious of random or stratified random sampling approaches to case-selection. This, of course, is not what the authors intend: the case of Algeria is offered as an example of a much larger sample of cases—25 in all. But, for heuristic purposes, let us discuss a few of these potential difficulties.

Begin by stratifying the total population of potential country cases ($N \approx 180$) across three dimensions that are deemed to be theoretically relevant, e.g., *socioeconomic status* (rich/poor), *civil war* (yes/no), and various *regions* including Africa, the Americas, Eurasia, and the Pacific. The intersection of these dimensions sub-divides the universe of country-cases into sixteen sub-strata. It should be obvious that a fairly large sample will be necessary in order to represent, in a plausible fashion, the full range of cases in the population—at a *minimum*, sixteen (one from each sub-stratum). However, this presumes a very high degree of homogeneity within each sub-stratum such that all cases within a sub-stratum yield virtually equivalent results (with respect to whatever causal proposition is being explored). In this setting, which we must imagine is extremely rare in the social sciences, it hardly matters how one chooses cases—random or purposive. Every selection procedure building on the aforesaid stratification will achieve the same results. If, on the other hand, there is some theoretic-

cally relevant variation across cases within a substratum, then a much higher number of cases will be necessary to produce a sample that can claim to be representative (probabilistically) of the broader population of nation-states. Clearly, we are in large- N territory.

Now, if the researcher sacrifices the goal of representativeness she is free to choose among particular substrata, ignoring others. (Note that this is not what survey researchers mean by *over-sampling* since it is no longer possible to reconstruct, by weighting, a truly representative sample.) Suppose she decides, on some basis, that she will choose cases only from the substratum of cases that are poor, civil-war prone, and Asian. In principle, this could be handled by random draw, removing the researcher from the decision and any potential bias that might result. Yet, there are several reasons to think twice about this procedure.

First, there is the problem of sampling variance within each substratum—presumably less than one would find in the population at large (since the substratum is chosen with an eye to creating greater homogeneity), but still perhaps enough to give pause. China and Burma, both of whom qualify as members of the identified substratum, are very different places. Thus, representativeness of the substratum (let alone of the larger population) is unlikely to be achieved in an $N=1$ sample chosen in this manner. Arguably, representativeness (in this limited sense) is more likely to be achieved when the researcher chooses the case purposefully than when she chooses randomly, for she can incorporate background knowledge of the cases (factors not included in the stratification) and exercise judgment about which case lies nearest to the mean along relevant dimensions.

The second obstacle concerns the leverage (for causal inference) that a given case is likely to provide. Recall that a good sample is not only representative but also insightful. This can mean lots of different things, but to simplify let's say that some cases look more like natural experiments than others. Suppose that the civil war in China is accompanied by all sorts of theoretically extraneous factors (connected with the communist insurgency or foreign intervention) that have little to do with the theoretical hypotheses at hand. Burma, by contrast, has few of these potential confounders. Under the circumstance, Burma is clearly a better choice (all other things being equal). This, too, validates a purposive (non-random) approach to case-selection.

Third, there are practical factors like the language capacities of the researcher, the availability of documents and other sources, and the ability to gain entry to a country for field research. Again, one may find strong reasons to favor Burma over China (or China over Burma)—reasons that would be washed away if the case were chosen randomly.

To be sure, there is no limit, in principle, to the number of features that can be included in a stratification procedure. Every "purposive" feature mentioned above can be incorporated into the formal stratification, removing it from possible investigator bias. This also enhances the clarity and explicitness of the case-selection procedure. Yet, one wonders whether all of these factors can really be measured, *ex ante*, across the

entire population. And if they could, the resulting stratification would include so many dimensions that—with a fixed and moderate-sized population—each sub-stratum would be miniscule, allowing little scope for random draws.

In short, it is difficult to justify selecting a sample of *one or several* cases in a purely random or stratified random fashion. One can see why case study researchers want to attach proper names to their potential cases before finalizing their selection. Case knowledge is often revelatory.

Thus, random sampling within pre-identified sub-strata is sometimes viable (it is indeed suggested in my chapter as part of the "diverse-case" procedure). Where it is, it offers a potential solution to problems of researcher bias that might otherwise influence the case-selection procedure, a concern raised by Fearon and Laitin (see also Gerring 2007a: 145). Where it is not, the researcher must fall back on purposive (non-random) procedures. Either way, the most critical question is probably this: *on what principles* should the case-selection criteria—stratified or not—occur? How, in the context of a stratification, should the strata and sub-strata be selected? This is the main topic addressed in my chapter. (Note that the selection of strata, and the choice of sub-strata to over-sample, presupposes a selection principle for the strata; this I consider to be the purposive, or intentional, element.)

I shall now proceed to address some of the criticisms raised by Freedman's commentary. Freedman's summary comment on my chapter is that "some of the methods that Gerring proposes are ill-suited to qualitative research." I gather that his hesitancy about applying statistical methods to case study research stems, in part, from his view of the inherent limits of quantitative techniques when faced with nonexperimental data. I share this skepticism. Even so, it seems clear that case study research is often compelled to labor with nonexperimental data. This being the case, the relevant question is whether problems of causal attribution endemic to observational studies are made any worse by the application of statistical models to aid in the process of case selection. Here, I am agnostic.

Before continuing, I should clarify that I am emphatically not proposing that statistical analysis be applied to very small samples. This, combined with the ubiquitous assignment problem posed by observational data, is a recipe for disaster. What I am proposing is that quantitative techniques might find employment, at least on certain occasions, in the selection of cases for in-depth research focused on one or several cases. (I would also argue that there is no reason to preclude the statistical analysis of large-sample *within-case* evidence—though this issue was not raised in the chapter.)

I am assuming, of course, that the population of interest is large enough to justify the employment of regression, matching, and other suggested techniques. I write: "In certain circumstances, the case-selection procedure may be structured by a quantitative analysis of the larger population" (646). Again, "Sometimes, these principles can be applied in a quantitative framework and sometimes they are limited to a qualitative framework. In either case, the logic of case selection remains quite similar, whether practiced in small- N or large- N contexts" (*ibid.*).

Note that all of these techniques are introduced as techniques of case *selection* (as signaled in the title of the chapter), not case *analysis*. The idea is that, in some instances, the relevant requirements for a statistical analysis may be met, and in these instances it makes sense to formalize the case-selection strategies that would otherwise be carried out in a qualitative manner. Thus, with respect to the “deviant-case” strategy, rather than choosing a case that appears unusual with respect to some informal model of causal relationships, one might actually test these assumptions formally in a statistical model, choosing a case (or cases) with a high residual(s).

Occasionally, the goal of a case study is to confirm/disconfirm a statistical model. Here, an appropriate strategy of case selection might be an “influential-case” analysis—where relevant cases are identified by examining hat matrix and Cook’s distance statistics for individual cases (developed initially in Seawright and Gerring 2008). Similarly, other quantitative techniques such as cross-tabulations may be helpful in the context of a “diverse-case” analysis. To say that a case-selection procedure is purposive does not, therefore, imply that it must be small-*N* or qualitative.

Of course, it is always an open question how much confidence one ought to place in large-*N* statistical models. Yet, an article on case selection in case study analysis did not seem the appropriate venue to expatiate on a point that so many others (notably David Freedman) have persuasively argued. I stated explicitly that “relevant data must be available for [a] population...on key variables, and the researcher must feel reasonably confident in the accuracy and conceptual validity of these variables. [Further,] all the standard assumptions of statistical research (e.g., identification, specification, robustness) must be carefully considered, and wherever possible, tested.” I then warned against “the unreflective use of statistical techniques.” Doubtless, some people will continue to use statistical techniques even where they are not warranted. But this potentiality should not prevent us from discussing instances in which the employment of statistical techniques is warranted.

At this point, it may be helpful to formally define the key term, “case study.” Sometimes, the case study method is equated with qualitative methods. My understanding of the concept is different (Gerring 2007a, 2007b). A case study, as I see it, is most usefully defined as the intensive study of a single case, or several cases, where the purpose is to shed light on a broader population of cases. It should not be equated with qualitative methods, since we already have a term for this concept. To be sure, any *cross-case* analysis—i.e., in a Millean-style comparative study—would have to be qualitative, for the sample is extremely small (by definition). But case selection and within-case analysis may be qualitative and/or quantitative. (The resulting two-by-two matrix produces four cells, and all are occupied.)

I turn now to the comparison of case study techniques and experimental techniques—a minor point of the chapter but a central objective in previous work (Gerring 2007a: chapter 6; Gerring and McDermott 2007). I concur with Freedman that “The logic of randomized controlled experiments is...

central, even for observational research.” Thus, in the course of discussion of Mill’s most-similar method (aka the “method of difference”) I make several analogies to experimental methods.

For example, in order to reduce background noise, experimentalists often stratify a sample into relatively homogeneous (“most-similar”) sub-strata prior to treatment. The treatment is then randomized *within* each sub-strata (or block). If the sub-strata are very small, e.g., blocks of two, the procedure may be described as an iterated most-similar comparison with randomized treatment. In this respect, I think it is fair to say that both Millean and experimental studies often attempt to identify samples (or sub-samples) that are as internally homogeneous as possible. I regret that I did not offer some explication of this point—which is not self-evident—in the chapter.

A great deal of ground has been covered in this all-too-brief review. The common thread running through the narrative is that our understanding of case study research is enhanced when we make comparisons and contrasts with other sorts of research—e.g., with large-*N* cross-case analysis with observational data or with experimental research. Sometimes, techniques not usually associated with the case study can be utilized in the selection of a few cases for intensive analysis. Sometimes, on the other hand, these techniques are inappropriate. I trust that we are moving closer to a consensus on these matters.

In any case, there are many hopeful signs that the gulf that has traditionally separated case study and non-case study methods is narrowing. This colloquy is an excellent example of that propitious development. Let me close by thanking David Freedman, James Fearon, and David Laitin for facilitating this exchange. I only wish I had the benefit of their counsel when crafting earlier projects.

Note

¹ All direct quotations are from my chapter in the Oxford volume, unless otherwise noted.

References

- Eckstein, Harry. 1975. “Case Studies and Theory in Political Science.” In *Handbook of Political Science, vol. 7. Political Science: Scope and Theory*. Fred I. Greenstein and Nelson W. Polsby, eds. (Reading, MA: Addison-Wesley), 94–137.
- Gerring, John. 2007a. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Gerring, John. 2007b. “The Case Study: What it is and What it Does.” In *Oxford Handbook of Comparative Politics*. Carles Boix and Susan Stokes, eds. (New York: Oxford University Press), 90–122.
- Gerring, John and Rose McDermott. 2007. “An Experimental Template for Case-Study Research.” *American Journal of Political Science* 51:3 (July), 688–701.
- Seawright, Jason and John Gerring. 2008. “Case-Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options.” *Political Research Quarterly* 61:2 (June), 294–308.

Choosing Cases for Case Studies: A Qualitative Logic

Gary Goertz

University of Arizona
ggoertz@u.arizona.edu

David Freedman in his essay has called into question some of the advice given by Fearon-Laitin and Gerring in their chapters in the *Oxford Handbook of Political Methodology*. I would like to second those criticisms and extend them in various ways. In particular, Freedman does not address in much detail the regression or logit model which underlies, explicitly or implicitly, both these chapters—and more generally Gerring’s book on case study methodology (2007). This “regression approach” to case studies (to give it a name) informs much discussion about case studies and qualitative methods, going back to King, Keohane, and Verba (1994) and more recent works such as Lieberman (2005). In these few pages I can but sketch a rationale for choosing cases following a different logic of research. In the first part of the essay I address the choice of case studies from a qualitative logic of research. In the second part, I briefly describe a “descriptive-causal” approach to case study selection which is different from the regression logic of Fearon, Laitin, and Gerring.

Case Studies Selection and Research Agendas

Many, if not most, research projects start with a decision to explain some phenomenon, $Y = 1$ for short. This can be war, democracy, voting, or whatever. As such one will naturally select cases of $Y = 1$ for intensive scrutiny, because explaining $Y = 1$ is exactly the overall goal of the research project.

With so much focus in methodology classes on the problems with “selecting on the dependent variable” it is easy to lose sight of the key methodological principle that one should select some individual cases of $Y = 1$ for intense examination. For example, when a new disease occurs medical researchers first focus on people with the disease in order to understand it. When AIDS was first discovered there was intense concentration on those who had the disease. Ragin is one of the few who has constantly stressed the importance of focusing on the $Y = 1$ cases (1987, 2000, 2008). A first principle of case selection of case studies is:

Principle 1: One should choose, *diversely*, among the *good* instances of $Y = 1$ for case studies.

According to this principle you want to choose a *diverse*, not random, set of cases because you do not want to miss an important causal path to Y . Fearon and Laitin stress the value of random selection. Both they and Freedman recognize the pitfalls of “cherry-picking” and that random selection can help avoid this problem. At the same time Fearon-Laitin want to use random selection to choose “representative” or “typical” cases. Because they start with a regression model, they work

from the presumption that there is a β_i that is the typical causal effect of X_i . A qualitative research logic is much more likely to start with an INUS model of causation where there are multiple paths to Y , such as $Y = x * A * B + X * B * C$ (lower case means absence of factor). If you start with an INUS view of the world you do not necessarily believe there is one representative causal effect of X , since, depending on the path, the presence of X or its absence x may be a cause of Y . So one looks for diverse cases in order to not miss causal paths. In short, whenever Gerring, Fearon, and Laitin use the idea “random” I would suggest replacing it with “diverse.”¹

Principle 1 also stresses that one should choose among “good,” i.e., unambiguous, cases of $Y = 1$. In the case of civil wars this means one should not choose randomly among all cases of $Y = 1$. There is certainly a good percentage of cases of civil war that are marginal, or “gray” cases of civil war. Definitions of phenomena draw boundaries; almost inevitably there are cases near those boundaries that may not fit the concept very well. If one is beginning to study AIDS, one would not choose cases that may or may not be AIDS or cases that seem atypical of AIDS.

After one has expanded some significant shoe leather understanding deeply some $Y = 1$ cases one might begin to have some ideas about the causes of Y . The concerns with cherry-picking are real and I think they are most often tied to selecting cases based on X , particularly $X = 1$. It is natural to focus on cases where the author’s theory works. Hence from a qualitative point of view the risks of bias are in some sense larger when selecting on X than selecting on Y .

These potential causes, $X = 1$, then lead to choosing cases that allow one to see how plausible (to use Eckstein’s terminology) X is as a potential cause of Y . Thus we will choose cases where X is present to see if we can figure out the causal mechanism linking X to Y . Typically, we will focus our attention on the (1,1) cases, i.e., $Y = 1$ and $X = 1$ cases, because these allow for causal process tracing (George and Bennett 2005). Here we have a second principle for choosing instances for case studies:

Principle 2: Select, *diversely*, cases of $X = 1$.

For example, researchers trying to explain lung disease noticed that this seemed to be common among smokers. This naturally led then to efforts such as experiments on rats where they were forced to smoke a lot. The rationale for diversity in Principle 2 is the same as in Principle 1: one wants to detect multiple causal paths to $Y = 1$.

It cannot be stressed enough that part of case study methodology involves counterfactual analysis (Tetlock and Belkin 1996; see Levy 2008 for an excellent discussion). Counterfactual methodology turns $X = 1$ into $X = 0$. This is in part why we can focus on the $X = 1$ cases, because we later turn them into $X = 0$ via counterfactual analysis.

For some X s it may not be possible, or may be very difficult, to do a good counterfactual analysis. If we are exploring the impact of wealth or GDP/capita on civil war, it would be hard to say what is the likelihood of civil war if, for example, Sierra Leone were a wealthy, developed country.² The fact that

this is a difficult counterfactual (see Ragin 2008 for this concept) means that selecting a case study of $X = 0$ based on wealth will be problematic as well. This is related to the well-known “minimum rewrite rule” for counterfactuals which recommends only modest changes in X for counterfactuals. This is so important that I think it needs to be elevated to a principle:

Principle 3: If the counterfactual $X = 0$ is problematic in individual cases of $Y = 1, X = 1$, then choosing actual instances of $X = 0$ for case study analysis is problematic as well.

Principles 1 and 2 stress that we choose case studies based on sampling on $Y = 1$ or $X = 1$. This leads to a principle which is a corollary of the first two:

Principle 4: Cases of $Y = 0$ and $X = 0$ are often not very useful for intensive case study examination.

Jim Mahoney and I have made the argument elsewhere (2004) that the (0,0) cases are problematic for qualitative researchers. Often there is a large number of these cases. Random selection among this (often very large) number is unlikely to produce cases that will be useful for a case study. These cases might be very important in a large- N statistical analysis, but much less useful for a case study. For example, it is crucial to the large- N study of the linkage between smoking and lung cancer to include (0,0) cases.

I suggest that these principles for choosing instances for case study research are what researchers often naturally do, and at the same time good practice. Underlying the belief that multi-method research is good lies the notion that different methods give us different information and different views of the phenomenon. If one bases case study research on the regression model we lose the distinctive, and I think complementary, advantages of case study methods vis-à-vis regression methods.

The Descriptive–Causal Approach to Case Study Selection

Fearon-Laitin and Gerring explicitly link the selection of case studies to regression or logit models. In this section I propose that we can and should use our knowledge of the cases and of patterns in the cases to restrict our attention to quite limited regions of the data (or absence of data) determined by our empirical and theoretical interests. This I call, for reasons that should become clear, the “descriptive–causal” approach to case study selection.

For purposes of contrast I will use Gerring’s (2007) example of the relationship between wealth, aka GDP/capita, and democracy, which is a central example in his core Chapter 5, and more generally in the literature on the social, economic, and political requisites and correlates of democracy.

The descriptive–causal approach takes advantage of the fact that we have often accumulated some basic descriptive knowledge about the cases and about some general patterns in the data. Selection of case studies then relies on case knowledge. I use the term descriptive–causal to apply to descriptive statistics that have causal implications. For example, the demo-

cratic peace can be formulated as “Democracies do not fight wars with each other.” This is *descriptive* in the sense that it gives the frequency of occurrence of a phenomenon. It is *causal* in the sense that it makes a link between a potential cause, democracy, and a potential effect, war.

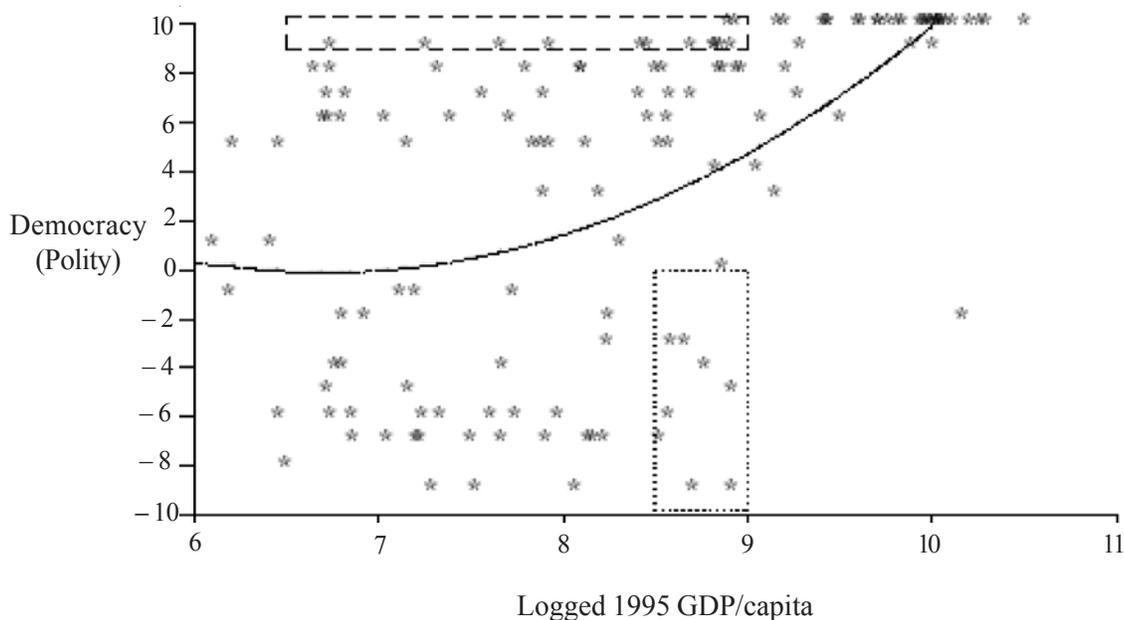
Lipset’s American Sociological Association presidential address (1994) is in fact a review of the literature that he was so instrumental in launching several decades earlier. This review is full of descriptive–causal patterns. For example, “every country with a population of at least 1 million that has emerged from colonial rule and has had a continuous democratic experience is a former British colony” (Lipset 1994: 5). Dahl (1971) says that “all highest-level [developed] countries are polyarchies” (cited in Diamond 1992: 97). Such examples can easily be multiplied.

The key thing about these descriptive–causal statistics is that they point to regions of the data where we might want to focus our attention in the form of an intensive case study. Figure 1 (adapted from Gerring 2007) illustrates the descriptive–causal approach and how it differs from a regression one. As a point of reference I have drawn a regression curve (OLS) through these data (not in the original Gerring figure). Take for example Przeworski et al.’s well-known analysis (2000) of the wealth–democracy relationship which depends on, roughly, noting that (1) most highly developed countries are democracies and (2) such countries rarely fall back into authoritarian systems. If these descriptive–causal patterns are correct then over time we should see very few cases in the lower-right hand corner of figure 1. Thus the interesting thing is not the regression curve through the middle of the data, but rather the lower-right-hand region. Given our interest in this region we might explore the one outlier in the figure (Singapore). It seems that this pattern really starts at a logged GDP/capita of about 9. For countries that are poorer the pattern does not seem to hold. Hence we might choose a case study in the region delimited by the dotted lines.

Another well-known descriptive–causal pattern in the literature is that it seems very difficult for poor countries to become high-quality stable democracies. Much of the early work focused on this particular pattern. This pattern directs our attention to the upper left and center part of figure 1, where we should see poor democracies if they exist. If we consider 10 on the polity scale to be “high-quality democracy” then we notice that this particular causal effect does not kick in until a logged GDP/capita of about 8.5. Here too we want to choose some case studies in the zone bounded by the dashed lines to explore more closely this potential causal relationship.

As figure 1 makes clear, if we are interested in these descriptive–causal patterns, working from a regression or probit line is of little use. If we randomly choose cases or choose cases based on the regression, the likelihood that they would be informative for our causal purposes is very low.

This essay does not say that the regression approach to case studies is not useful. Rather it argues that there are alternative ways to think about selecting case studies. In a different empirical or theoretical setting the regression approach may just be what the doctor ordered. At the same time much of

Figure 1: Descriptive-Causal Patterns: Wealth and Democracy

Source: Adapted from Gerring 2007: 96.

empirical, case-oriented knowledge that is so important to qualitative scholars is expressed in descriptive-causal claims like those of Lipset, Diamond, and Dahl. It is not surprising that these claims lie at the core of Ragin's Boolean and fuzzy set methodologies. They are a means of formalization of many descriptive-causal claims.

Using Case Studies to Evaluate Scope Conditions

With the notable exception of Dul and Hak (2008), the literature on case studies has not seen them as useful in exploring scope conditions. As Freedman notes in his essay, most statistical data analyses are based on samples or on populations of convenience. Mahoney and I have argued (2006) that qualitative scholars are often more concerned with scope conditions because they are much more concerned with cases that do not fit the theory.

Figure 1 illustrates nicely how one can use case studies to evaluate scope conditions. In figure 1 Gerring and Seawright have left out a handful of cases, which in fact lie in the lower-right corner, and hence are of particular interest to the Przeworski et al. pattern, and can be potentially seen as problematic.

If we look at these countries a new pattern becomes very clear: they are all wealthy countries because of large oil reserves, such as Saudi Arabia. The comparative politics literature (e.g., Ulfelder 2007) has already noticed that these countries suffer from the "natural resource curse." These authoritarian governments can remain in power without taxing their subjects. Hence it might be very reasonable for Gerring to exclude these cases from consideration because they do not fit the causal mechanisms that we find in the rest of the world.

While these cases are outliers in the regression model, a random selection of outliers would never detect this pattern

(or more precisely, would detect it with extremely low probability). There are lots of outliers from the regression curve. It is because we have particular theoretical and empirical interest in the lower right-hand corner that these outliers become important to us and an analysis of the individual cases can lead to scope restrictions.

The key point here is that cases are not representative of some given population but rather that the population is constructed via of knowledge of the cases, cases studies, and our causal analyses of them.

Conclusion

In this essay I have argued that we should, and usually do, have clear purposes for selecting instances for intensive case study. In general, we normally want to focus on the cases where $X=1$ and $Y=1$; we probably want to avoid cases of (0,0) unless we have a clear substantive rationale.

Qualitative scholars also select cases based on their knowledge of the cases and patterns in the cases. This descriptive-causal knowledge can point to particular regions of the data for selection of case studies.

One likewise uses case studies to construct and delimit populations. Instead of being given or taken by convenience, qualitative scholars construct populations using their knowledge of the cases.

I am a firm believer in the toolbox metaphor for methods. I see the descriptive-causal approach to case studies as another useful tool. Depending on the empirical and theoretical goals it might be more appropriate than the regression approach and in some circumstances it might be less.

Notes

¹Of course one needs to define the dimensions of diversity; one option is region as in the Fearon-Laitin chapter.

²This leads to King and Zeng's (2007) counterfactual critique of this literature.

References

- Diamond, Larry. 1992. "Economic Development and Democracy Reconsidered." In Gary Marks and Larry Diamond, eds. *Reexamining Democracy: Essays in Honor of Seymour Martin Lipset*. (Newbury Park, CA: Sage Publications).
- Dul, Jan and Tony Hak. 2008. *Case Study Methodology in Business Research*. Amsterdam: Elsevier.
- Fearon, James and David Laitin. 2008. "Integrating Qualitative and Quantitative Methods." In Janet Box-Steffensmier, Henry Brady, and David Collier, eds. *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press.
- George, Alexander and Andrew Bennett. 2005. *Case Studies and Theory Development*. Cambridge: MIT Press.
- Gerring, John. 2007. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Gerring, John. 2008. "Case Selection for Case Study Analysis: Qualitative and Quantitative Techniques." In Janet Box-Steffensmier, Henry Brady, and David Collier, eds. *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press.
- King, Gary, Robert Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- King, Gary and Langche Zeng. 2007. "When Can History be our Guide? The Pitfalls of Counterfactual Inference." *International Studies Quarterly* 51:183–210.
- Levy, Jack S. 2008. "Counterfactuals and Case Studies." In Janet Box-Steffensmier, Henry Brady, and David Collier, eds. *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press.
- Lieberman, Evan. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99: 435–52.
- Lipset, Seymour Martin. 1994. "Social Requisites of Democracy Revisited." *American Sociological Review* 59: 1–22.
- Mahoney, James and Gary Goertz. 2004. "The Possibility Principle: Choosing Negative Cases in Comparative Research." *American Political Science Review* 98: 653–69.
- Mahoney, James and Gary Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14: 227–49.
- Przeworski, Adam, et al. 2000. *Democracy and Development: Political Institutions and Well-being in the World, 1950–1990*. Cambridge: Cambridge University Press.
- Ragin, Charles. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- Ragin, Charles. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- Ragin, Charles. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.
- Tetlock, Philip E. and Aaron Belkin, eds. 1996. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton: Princeton University Press.
- Ulfelder, Jay. 2007. "Natural Resource Wealth and the Survival of Autocracy." *Comparative Political Studies* 40: 995–1018.

Rejoinder

David A. Freedman

University of California, Berkeley

I would like to begin by thanking James Fearon, David Laitin, John Gerring, and Gary Goertz for their comments, which help to clarify the issues on the table. Whatever differences remain, we can all agree that David Collier did a great job in organizing this discussion.

Fearon and Laitin

As I read the paper, Fearon and Laitin (2008) made a clear statement that case studies were a poor way to establish empirical regularities; large-*N* methods were to be preferred. The claim was justified using a narrow definition of "empirical regularities," which excluded pretty much everything except summary statistics (means, standard deviations, regression coefficients...). The paper segued to an implication that there was one natural way of integrating qualitative and quantitative research methods, with the former as ancillaries to the latter.

However, the story has a happy ending. Fearon and Laitin explain that my reading of the paper was not the intended reading. As it turns out, we agree on the following points. (i) There many ways to do good science. (ii) In particular, neither cluster of methods has a general advantage over the other. (iii) Therefore, there are many fruitful ways for qualitative and quantitative researchers to interact. (iv) When it comes to making causal inferences, case studies often have considerable power—although, to be sure, for statistical inference, bigger *N* is usually better.¹

There was also some back-and-forth about sampling. On this topic too, there is now reasonable agreement. As with other choices to be made in research, much depends on context and on background knowledge. Fearon and Laitin raise a new point, contrasting the messy realities of research design with "the pristine hypothesis-testing scenario assumed in statistics textbooks." This shaft is well-aimed, although the environment is target-rich: there are econometrics textbooks, psychometrics textbooks....

What can be said about the substantive research in Fearon and Laitin (2008)? I believe that Fearon and Laitin used their logit model descriptively, to find patterns in the data that suggest one causal theory or contradict another theory. They did not rely on the model to make causal inferences. Instead, they used the case studies to do the heavy lifting. They chose cases using an elegant and impartial technique. These are useful ideas, which should find many applications.

Gerring

I think there is agreement on the following points:

(i) Large-*N* researchers should use random samples, and often they do. Often, however, there is a divergence between the ideal and the real.

(ii) Modeling and matching are large-*N* techniques that

make stringent assumptions about data-generating mechanisms. These techniques can help us choose a small number of cases for in-depth study when (a) we are choosing those cases from a big, well-defined population, (b) there are complete data for all the cases in the population, and (c) the assumptions behind the modeling and matching are viewed as reasonable for the population. These conditions are clear in Gerring’s response, as they are in Seawright and Gerring (2008). They will seldom (if ever?) be satisfied in small-*N* research: even the first two might be problematic.²

(iii) If the experiment was done well, few assumptions are needed to analyze the data. Blocking subjects to achieve greater homogeneity may be a good idea, but that is something you do before randomization, not when you are analyzing the data.

(iv) Causal inferences are frequently based on observational studies rather than experiments, with elaborate modeling and matching to control for confounders. Assumptions play a large role in these proceedings, and the opportunity for error is correspondingly large. This makes a striking contrast with experiments. I would add, however, that in many cases our causal knowledge derives from well-designed observational studies, where the data do not require complex statistical analysis (Freedman 2005: Chapter 1); I think Gerring will agree.

Although this topic is only tangential to Gerring’s work, more should be said about non-response. Non-response rates are high for many surveys, and the level is generally rising. Even if we start from a probability sample, the actual respondents going into the analysis can be a lot like a convenience sample, because non-respondents and respondents can be very different (Freedman et al. 2007: Chapter 19).

To illustrate the magnitude of the problem, I will use three of the papers reprinted in Freedman (2005).³ These papers start with large probability samples. However, 50% to 75% of the data are missing, because subjects refused to cooperate with the survey, or declined to provide some of the data that were needed. This is especially poignant because the papers are grappling with the endogeneity of selection into treatments of one kind or another. However, endogeneity of selection into the sample is politely ignored. As Gerring (2008: 678) says, “Not all twists and turns on the meandering trail of truth can be anticipated.”

Goertz

I disagree with half of what Goertz says, but will only respond to three things: (i) the interpretation of Fearon-Laitin and Gerring, (ii) the philosophy of case selection, and (iii) the advice to ignore a cell in the 2 x 2 table.

(i) According to Goertz, “Fearon-Laitin and Gerring explicitly link the selection of case studies to regression or logit models.” I cannot see any explicit statement either in Fearon and Laitin (2008) or in Gerring (2008) to that effect. Of course, Fearon and Laitin are selecting cases in the context of a logit model. However, these scholars do not rely on the assumptions behind the model (Freedman 2005: Chapter 6) when selecting cases. Indeed, the principal recommendation on case

selection is to use stratified random samples. This has little to do with models.

What about Gerring? To be sure, a few of his techniques for small-*N* case selection are linked to regression models. In my view, previously noted, these suggestions will rarely be helpful. By contrast, most of his discussion—for instance, of typical and diverse cases—is blessedly model-free and generally useful. (Is this causation or just association?)

(ii) Goertz says, “The key point here is that cases are not representative of some given population but rather that the population is constructed via knowledge of the cases, case studies, and our causal analyses of them.” The statement comes perilously close to a recommendation that we should start with a theory, choose cases in conformity with that theory, and then conclude that the evidence supports the theory. No one is immune from this tendency, but it is a habit to be discouraged rather than encouraged.

(iii) Goertz considers a binary causal variable *X*, where *X* = 1 means the causal factor is present, while *X* = 0 means the factor is absent. There is a binary response variable *Y*. The data can be presented in a 2 x 2 table:

	<i>X</i>	
<i>Y</i>	1	0
1	A	B
0	C	D

Goertz recommends in favor of looking at cell A when doing qualitative research; he recommends against looking at cell D. Curiously, he adds that for large-*N* research, cell D “*might* be very important [emphasis supplied].”

Playing favorites with cells is a risky business. At least in my experience, it is often hard to see where the cases go until you study them. Moreover, despite Goertz’s reservations, all four cells are important to large-*N* researchers. Indeed, consider data like the following:

	<i>X</i>	
<i>Y</i>	1	0
1	10	20
0	20	??

If the number of cases in cell D is above 40, there is positive association; if the number is below 40, there is negative association. Since there is a fundamental difference between “*X* causes *Y*” and “*X* prevents *Y*,” cell D matters in large-*N* research, along with the other cells.⁴

The boundary between large *N* and small is salient in the present context. For the moment, let us set the boundary at *N* = 17. If you accept Goertz’s position, cell D can be relevant when *N* is above 17; cell D cannot be relevant when *N* is below 17. This is not a tenable position, and moving the boundary will not solve the problem.

If all four cells are relevant, close inspection of cases in all four cells has to be a good idea, at least under some circumstances. For example, a critic might assert that cases in cell D exhibit causal heterogeneity. The most straightforward way to rule that out is to look at cases in cell D.

The present exchange offers a real example. Fearon and

Laitin found cell D to be informative. This contradicts Goertz's position. In qualitative research, to be sure, examining only one cell in the table may sometimes be a good idea.⁵ However, advice that cell D should generally be ignored is, well, advice that should be ignored.

Conclusion

I will draw an empirical conclusion⁶ from the discussion: there are few recipes for good research. (Cooking and scholarship depend on somewhat different skill sets.) Nearly 50 years ago, my friend Larry Jackson defined the scientific method as "guess and verify." The only improvement I can make is to emphasize part of the recommendation: guess and verify.

Notes

¹ Fearon and Laitin point out that you do not want to increase N by stretching concepts. From my perspective, increasing sample size should reduce sampling error; but the effect on non-sampling error is unpredictable. Moreover, large- N research is often needed to demonstrate causation. Epidemiologic studies on the health effects of smoking, mentioned by Goertz, illustrate the point. For a brief review, see Freedman (1999).

² Fearon and Laitin are drawing a sample from a large, well-defined population to which a model has been fitted, so the first two conditions are satisfied. However, as indicated above, far from relying on the model, Fearon and Laitin are using the sample cases to test the model. Gerring indicates that the hat matrix and Cook's distance may be helpful in such contexts. I agree, but this favorable conjunction of circumstances is rare in qualitative research; in multi-method research, the story may be different.

³ In the fourth paper, the unit of analysis is the state, so the issues are a little different.

⁴ This discussion ignores sampling error, which is reasonable if N is large. The "odds ratio" is used to summarize the data, as is standard in epidemiology. Let a denote the number of elements in cell A, and so forth. If there are cases in all four cells, the odds ratio is $(a/c)/(b/d) = (a/b)/(c/d) = (ad)/(bc)$. The association is positive when the odds

ratio is above 1.0; the association is negative when the odds ratio is below 1.0. You need all four numbers to compute the odds ratio.

If I denotes the odds ratio, the causal interpretation is this: setting X to 1 rather than 0 multiplies the odds that $Y = 1$, by the factor I . Equivalently, if $Y = 1$ rather than 0, the odds that $X = 1$ are multiplied by the factor I . For additional information, see Gordis (2008).

⁵ Great work can be done with one cell, or even one case. Isn't de Tocqueville's *Democracy in America* a classic example of within-case analysis?

⁶ Of course, like others discussed earlier, this conclusion depends in part on context and background knowledge.

References

- Box-Steffensmeier, Janet M., Henry E. Brady, and David Collier. 2008. *The Oxford Handbook of Political Methodology*. Oxford University Press.
- Fearon, James and David Laitin. 2008. "Integrating Qualitative and Quantitative Methods." In Janet Box-Steffensmier, Henry Brady, and David Collier, eds. *The Oxford Handbook of Political Methodology* (Oxford: Oxford University Press), 756–76.
- Freedman, David A. 1999. "From Association to Causation: Some Remarks on the History of Statistics." *Statistical Science* 14: 243–58. Reprinted in *Journal de la Société Française de Statistique* 40 (1999), 5–32 and in John Panaretos, ed. 2003. *Stochastic Musings: Perspectives from the Pioneers of the Late 20th Century*. Lawrence Erlbaum Associates, 45–71.
- Freedman, David A. 2005. *Statistical Models: Theory and Practice*. Cambridge University Press.
- Freedman, David A., Robert Pisani, and Roger A. Purves. 2007. *Statistics*. 4th edition. New York: W. W. Norton & Company, Inc.
- Gerring, John. 2008. "Case Selection for Case Study Analysis: Qualitative and Quantitative Techniques." In Janet Box-Steffensmier, Henry Brady, and David Collier, eds. *The Oxford Handbook of Political Methodology* (Oxford: Oxford University Press), 645–84.
- Gordis, Leon. 2008. *Epidemiology*. 4th edition. Philadelphia: Elsevier-Saunders.
- Seawright, Jason and John Gerring. 2008. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61: 294–308.

Symposium: Field Experiments and Qualitative Methods

Natural and Field Experiments: The Role of Qualitative Methods

Thad Dunning

Yale University

thad.dunning@yale.edu

Political scientists increasingly use natural and field experiments in their research.¹ This raises the question—how do qualitative methods contribute to these research methodologies? I suggest here that there are strong complementarities between the use of such research designs and various kinds of qualitative methods. For example, case-based knowledge is often necessary to recognize and validate a potential natural experiment. The research skills associated with qualitative fieldwork, in turn, are often required for the implementation of field experiments. Qualitative methods can be crucial for designing experimental interventions, measuring outcomes, providing evidence on mechanisms, and even constructing random assignment mechanisms.

After discussing natural experiments from a variety of perspectives, I give a short example of how a field experiment may be used to explore the relationship between cross-cutting cleavages and ethnic voting in Mali, drawing on my recent joint research on this topic. As I describe, qualitative methods have contributed in both expected and unexpected ways to this project.

Natural Experiments and Qualitative Methods²

An illuminating if well-known exemplar of a successful natural experiment comes from John Snow's studies of cholera transmission (Freedman 1991, 1999, 2005; Dunning 2008). While its substantive domain lies far from the concerns of most social scientists, Snow's research illustrates the key role of qualitative methods in identifying and exploiting a natural experiment to make progress on an important problem.

Nineteenth-century London suffered a number of devastating cholera outbreaks. Although predominant theories linked cholera transmission to bad air (miasma) or to ground poisons, Snow became convinced that cholera was a waste- or water-borne infectious disease (Richardson 1887: xxxiv). In Snow's research, "causal process observations" (Collier, Brady, and Seawright 2004) were crucial, both for allowing Snow to formulate a hypothesis about the causes of cholera transmission and to provide evidence for the plausibility of this hypothesis. For example, Snow noted that outbreaks seemed to follow the "great tracks of human intercourse"; sailors who arrived in a cholera-infested port did not become infected until they disembarked, striking a blow to the miasma theory (Snow 1855: 2).

During London's cholera outbreak of 1853–54, Snow famously drew a map showing the addresses of deceased chol-

era victims. Because these addresses clustered around the Broad Street water pump in the Soho district, Snow argued that contaminated water supply from the pump caused the cholera outbreak. However, there were several anomalous cases: residences located near the pump where there had been no deaths from cholera, and residences far from the pump with cholera deaths. Snow used qualitative process tracing and a heavy dose of "shoe leather" (Freedman 1991) to probe these seemingly disconfirming outcomes (Snow 1855: 39–45). At a brewery located near the Broad Street pump, where cholera death rates were anomalously low, the proprietor told Snow that a fresh-water pump was installed on the premises—and that in any case the brewers tended to drink beer, not water (Snow 1855: 42). At another address, closer to another water pump than to Broad Street—and where there had been significant deaths from cholera—Snow learned that the deceased residents had preferred, for one reason or another, to take water at the Broad Street pump (Dunning 2008). Snow's experience as a clinician, his studies of the pathology of cholera deaths, and his spot map showing the proximity of victims to the Broad Street pump all provided bits of evidence, which suggested that cholera might indeed be an infectious disease carried by waste or water.

However, Snow's most powerful piece of evidence came from a natural experiment. Large areas of London were served by two water suppliers, the Lambeth company and the Southwark and Vauxhall company. Just prior to the cholera epidemic of 1853–54, the Lambeth company moved its intake pipe further upstream on the Thames, thereby "obtaining a supply of water quite free from the sewage of London" (Snow 1855: 68), while the Southwark and Vauxhall company left its intake pipe in place. After painstaking data collection, Snow constructed a simple cross-tab showing cholera death rates during the epidemic by source of water supply. For houses served by Southwark and Vauxhall, the death rate from cholera was 315 per 10,000; for houses served by Lambeth, it was a mere 37 per 10,000 (Snow 1855, Table IX, p. 86; presented in Freedman 2005).

Why did this constitute a credible natural experiment? Unlike true experiments, the data used in natural experiments come from naturally occurring phenomena—actually, in the social sciences, from phenomena that are often the product of social and political forces. Because the manipulation of the treatment, intervention, or independent variable is not generally under the control of the analyst, natural experiments are, in fact, observational studies. However, unlike other non-experimental approaches, a researcher exploiting a natural experiment can make a credible claim that the assignment of the non-experimental subjects to treatment and control conditions is "as-if" random. Outcomes are compared across treatment and control groups, and both *a priori* reasoning and empirical evidence are used to validate the assertion of randomization.

Thus, random or as-if random of assignment to treatment and control conditions—in Snow’s study, the water supply source—constitutes the defining feature of a natural experiment. This implies that at least as a necessary if not sufficient condition, the treatment and control groups are balanced with respect to other (measurable) variables that might explain cholera deaths. Notice that in a natural experiment, this is achieved not by statistical adjustment on the part of the analyst but rather by nature’s as-if randomization. Snow presented various sorts of evidence to establish this pre-treatment equivalence between the two groups. In his own words,

The mixing of the (water) supply is of the most intimate kind. The pipes of each Company go down all the streets, and into nearly all the courts and alleys. A few houses are supplied by one Company and a few by the other, according to the decision of the owner or occupier at that time when the Water Companies were in active competition. In many cases a single house has a supply different from that on either side. Each company supplies both rich and poor, both large houses and small; there is no difference either in the condition or occupation of the persons receiving the water of the different Companies... It is obvious that no experiment could have been devised which would more thoroughly test the effect of water supply on the progress of cholera than this (Snow 1855: 74–75).

Moreover, residents did not appear to self-select into their source of water supply: decisions regarding water companies were often taken by absentee landlords, the decision of the Lambeth company to move its intake pipe was taken before the cholera outbreak of 1853–54, and existing scientific knowledge did not clearly link water source to cholera risk. As Snow puts it, the pipe’s move meant that more than three hundred thousand people were:

divided into two groups *without their choice, and, in most cases, without their knowledge*; one group being supplied with water containing the sewage of London, and... the other group having water quite free from such impurity (Snow 1855: 75; emphasis added).

The cholera example provides several useful lessons about the elements of a successful natural experiment (see Freedman 1991, 1999). Snow went to great lengths to gather evidence and to use *a priori* reasoning to argue that only the water supply distinguished houses in the treatment group from those in the control group, and thus the impressive difference in death rates from cholera was due to the effect of the water supply. It is also worth noting that, while the natural experiment may have been the *coup de grace* in Snow’s painstaking investigation into the causes of cholera transmission, his use of this natural experiment was complemented and indeed motivated by the other evidence that he had gathered. The body of evidence Snow compiled depended on his detailed knowledge of the progress of previous cholera outbreaks in England, on his ability to cull information from a variety of sources, and especially on his willingness to do on-the-ground process tracing and close-range exploration of

seemingly disconfirming cases (Dunning 2008). This kind of close-range research also gave him the information he needed to discover and exploit his natural experiment, while his apparently innate sense of good research design led him to recognize the inferential power of the approach.

Social-Scientific Examples

Several of the elements of Snow’s successful natural experiment can be found in recent social-science applications, as well. Brady and McNulty (2004), for example, are interested in examining how the cost of voting affects turnout. In California’s special gubernatorial recall election of 2003, in which Arnold Schwarzenegger became governor, the elections supervisor in Los Angeles County consolidated the number of district voting precincts from 5,231 (in the 2002 regular gubernatorial election) to 1,885. For many voters, the physical distance from residence to polling place was increased, relative to the 2002 election; for others, it remained the same. Those voters whose distance to the voting booth changed—and who therefore presumably had higher costs of voting, relative to the 2002 election—constituted the treatment group, while the control group voted at the same polling place in both elections.

The consolidation of polling places in the 2003 election arguably provides a natural experiment for studying how the costs of voting affect turnout. A well-defined intervention, the closing of some polling places and not others, allows for a comparison of average turnout across treatment and control groups. The key question, of course, is whether assignment of voters to polling places in the 2003 election was as-if random with respect to other characteristics that affect their disposition to vote. In particular, did the county elections supervisor close some polling places and not others in ways that were correlated with potential turnout? Brady and McNulty (2004) raise the possibility that the answer to this question is yes, and indeed they find some evidence for a small lack of pre-treatment equivalence on observed covariates such as age across groups of voters who had their polling place changed (i.e., the treatment group) and those that did not. Thus, the assumption of as-if random assignment may not completely stand up either to Brady and McNulty’s careful data analysis or to *a priori* reasoning (elections supervisors, after all, may try to maximize turnout). Yet pre-treatment differences between the treatment and control groups are small, relative to the reduction in turnout associated with increased voting costs. After careful consideration of potential confounders, Brady and McNulty can convincingly argue that the costs of voting negatively influenced turnout, and a natural experimental approach plays a key role in their study.

Another increasingly common class of natural experiments exploits the existence of political or jurisdictional borders that separate similar populations of individuals, communities, firms, or other units of analysis, some exposed to a treatment or policy intervention and others not; in Dunning (2008), I review several studies and discuss the strengths and limitations of this form of natural experiments. Posner (2004), for example, studies the question of why cultural differences be-

tween the Chewa and Tumbuka ethnic groups are politically salient in Malawi but not in Zambia. Separated by an administrative boundary originally drawn by Cecil Rhodes' British South African Company and later reinforced by British colonialism, the Chewas and the Tumbukas on the Zambian side of the border are apparently identical to their counterparts in Malawi, in terms of allegedly objective cultural differences such as language, appearance, and so on. However, Posner finds very different inter-group attitudes in the two countries, with Chewas and Tumbukas in Malawi more likely to report an aversion to inter-group marriage and a disinclination to vote for members of the other group.

Posner argues convincingly that long-standing differences between Chewas and Tumbukas located on either side of the border cannot explain the very different inter-group relations in Malawi and in Zambia; a key claim is that "like many African borders, the one that separates Zambia and Malawi was drawn purely for [colonial] administrative purposes, with no attention to the distribution of groups on the ground" (Posner 2004: 530). Instead, the factors that make the cultural cleavage between Chewas and Tumbukas politically salient in Malawi but not in Zambia should presumably have something to do with exposure to a treatment (broadly conceived) on one side of the border but not on the other. Posner suggests that contrasts between inter-group attitudes of Chewas and Tumbukas in Malawi and Zambia are explained by the different sizes of these groups in each country, relative to the size of the national polities, which changes the dynamics of electoral competition and makes the groups political allies in Zambia but rivals in Malawi (see also Posner 2005).

Yet in order to argue this, Posner has to confront a key question which, in fact, sometimes confronts randomized controlled experiments as well: what, exactly, is the treatment? Or, put another way, which aspect of being in Zambia as opposed to Malawi causes the difference in political and cultural attitudes? Posner provides evidence that helps rule out the influence of electoral rules and the differential impact of missionaries on each side of the border. Rather, he suggests that in Zambia, Chewas and Tumbukas are politically mobilized as part of a coalition of Zambians living in the country's Eastern region, since alone neither group has the size to contribute a substantial support base in national elections, whereas in smaller Malawi (where each group makes up a much larger proportion of the population), Chewas are mobilized as Chewas and Tumbukas as Tumbukas (see also Posner 2005).

Clearly, the hypothesized intervention here is on a large scale—the counterfactual would involve, say, changing the size of Zambia while holding constant other factors that might affect the degree of animosity between Chewas and Tumbukas. This is quite different from imagining changing the company from whom one gets water in nineteenth-century London; one may question whether a manipulationist account of causation is most appropriate here (see Goldthorpe 2001 and Brady 2002). However, Posner's investigation of the plausibility of the relevant counterfactuals provides an example of "shoe leather" (that is, walking from house to house to find nuggets of evidence and rule out alternative explanations) in

the tradition of John Snow (Freedman 1991).

In natural experiments, a key question is whether treatment assignment really is as-if random, that is, independent of other factors that might explain differences in average outcomes across treatment and control groups. The assertion of as-if random assignment may be more compelling in some contexts than in others. As I discuss in Dunning (2008), it may be useful to conceptualize a "continuum of plausibility" that assignment to treatment and control is really as-if random; in that article, I place several recent studies along such a continuum and discuss ways in which the as-if random criterion may be partially validated with evidence as well as *a priori* reasoning (Dunning 2008).

For present purposes, the central point is simply that qualitative methods and case-based knowledge may play an important role in efforts to exploit as well as to validate natural experiments. Close knowledge of specific substantive domains may allow analysts to find and exploit credible natural experiments (see also Malesky, this symposium). And while simple quantitative techniques are also important for partially validating the claim of as-if random assignment (for example, for demonstrating equivalence on measured non-treatment variables across treatment and control groups), leveraging case-based knowledge about the substantive domain under investigation is also crucial to convincing applications of the natural-experimental approach.

Field Experiments and Qualitative Methods

In a randomized controlled experiment, subjects or units are randomized to treatment and control, and the intervention or manipulation is under the control of an experimental researcher (Freedman, Pisani, and Purves 1997). The main attraction of true (randomized controlled) experiments is that they solve pervasive problems of confounding and selection bias: random assignment ensures that treated and untreated groups are equivalent prior to the intervention, up to random error.³ With a large enough number of units, random error will play only a small role, and post-intervention differences across the treatment and control groups can be reliably attributed to the effect of treatment.

Field experiments—that is, randomized controlled experiments in which the "conditions under which a causal process of interest occurs are simulated as closely as possible" (Gerber and Green 2008)—offer many synergies with qualitative methods. As Gerber and Green (2008) point out, by definition, field experiments constitute "the conjunction of two methodological strategies, experimentation and field work." In some obvious ways, then, the skills associated with some qualitative researchers, particularly those who do fieldwork, are requisite for field experiments as well. The close case-based knowledge associated with some qualitative research may be vital for recognizing the opportunity to conduct a field experiment, and the social and networking skills often associated with qualitative fieldwork appear to be the *sine qua non* of many field experiments, as well.

Qualitative methods may play several other important roles in field experiments, however. Although not my main

focus here, one important potential contribution of qualitative methods is in identifying mechanisms, which is a crucial part of causal inference. For example, an experiment may allow the estimation of a causal effect without, however, illuminating the mechanism through which the cause produces its effect. Qualitative information may provide insights or information on context and mechanism, perhaps in the form of what Collier, Brady, and Seawright (2004) call “causal process observations.” (In addition, other experiments might be designed to elucidate the mechanism).

Yet there are also many other ways in which qualitative methods can contribute to field experiments, beyond simply field research skills. For example, they can help analysts confront challenges involved in measuring outcomes, designing treatments, recruiting participants, and even randomizing subjects to treatments. My objective in the rest of this article is to describe the contributions of qualitative methods to an ongoing experiment on ethnic politics in Mali. I first describe the experiment briefly, in order to set the stage for my discussion of qualitative methods.

Cross-Cutting Cleavages and Ethnic Politics: An Experiment in Mali

Social scientists often ascribe the absence or moderation of ethnic conflict to cross-cutting cleavages—that is, the presence of alternate dimensions of identity or interest, along which members of the same ethnic group may have diverse allegiances. Despite a rich theoretical literature, however, the empirical effects of cross-cutting cleavages are notoriously difficult to estimate. One goal of my ongoing research, conducted jointly with Yale undergraduate Lauren Harrison, is to formulate an experimental method for investigating the political effects of cross-cutting cleavages.

In Mali, despite substantial ethnic diversity, levels of ethnic conflict are persistently low. Unlike some Sub-Saharan countries, parties do not form along ethnic lines, and ethnicity is a poor predictor of individual vote choice. One set of explanations advanced for this African anomaly focuses on an informal institution called *cousinage* (loosely translated as “joking cousinship”). In Mali as well as in Sénégal, The Gambia, Guinea, western Burkina Faso, and the northern Ivory Coast—areas either formerly part of the Mali Empire (c. 1230–1600) or subject since to significant immigration from those areas—families historically formed alliances on the basis of patronyms. These historical alliances are now invoked in everyday social interactions. Today in Mali, for instance, if someone with the last name Keita meets someone named Coulibaly on the street, these two fictive cousins may invoke a standard set of jokes, even if they have never previously met. The jokes reinforce the social bonds understood to inhere in their relationship.

For our purposes, these alliances constitute cross-cutting cleavages, because they occur across as well as within ethnic groups.⁴ Despite a substantial literature on the alleged pacifying effects of *cousinage* (see Canut and Smith 2006; Davidheiser 2006: 837; Launay 2006; among early anthropologists, Mauss 1928 and Radcliffe-Brown 1940), it appears to us that this claim has not been subjected to empirical scrutiny

that would allow valid inferences about causal effects. We extend the hypothesis to explain not only the absence of ethnic conflict, generically, but also the apparent absence of ethnicity in electoral politics, asking why, in an ethnically-diverse African polity, ethnicity does not predict individual vote choice, and parties do not form along ethnic lines. Our extension of the cross-cutting cleavage (*cousinage*) hypothesis to explain political preferences and patterns of electoral competition in Mali is new and to our knowledge has not been previously tested.

We developed an experimental design to estimate the effects of *cousinage* relations on evaluations of political candidates and their speeches. First, we videotaped two Malian actors delivering the same speech, which focused on standard themes in Malian political campaigns; in initial field trials in the capital of Bamako, 56% percent of experimental subjects said the speech “reminded them of a speech they had heard on a previous occasion.” The speech was delivered in Bambara, which is the lingua franca of Bamako (and of Mali).⁵ We then recruited experimental subjects by canvassing all of Bamako’s neighborhoods (*quartiers*), approaching men and women sitting outside homes (or knocking on doors) and asking subjects if they would participate in a study on political speeches.⁶ We administered a screening questionnaire to each potential subject, asking for each subject’s first and last name and ethnic identity, along with various other personal information; this allowed us to assign subjects randomly to the treatment conditions, as described below.⁷ Experimental subjects then viewed our videotaped political speeches on a portable DVD player or laptop, using headphones.⁸ Finally, subjects then answered questions about the content of the speech and the politician who delivered it. For instance, they answered questions about the global quality of the speech, whether the speech made them want to vote for the candidate, and specific questions about candidate attributes such as competence, likeability, and intelligence.

The manipulation in this experiment consisted of what subjects were told about the politician’s last name. In Mali, last name conveys information about both ethnic identity and about *cousinage* ties. Thus, varying the politician’s last name allowed us to vary the treatment along two dimensions: the ethnic relationship of the politician and the subject (same ethnicity/different ethnicity) and their *cousinage* relationship (joking cousins/not joking cousins). Our resulting experimental design had six treatment conditions, four of which are shown in the cells of Table 1. We also added a fifth condition, in which the subject was provided with no information about the last name of the politician (and thus no information about ethnicity or *cousinage* ties), and a sixth treatment condition, in which the politician had the same last name as the subject.⁹

According to our hypotheses, a joking cousin relationship between voters and politicians should moderate the negative effect of ethnicity on voters’ evaluations of politicians. We expect evaluations of politicians to be more positive on average if the politician is a co-ethnic: thus, in Table 1, we expect to find that mean evaluations of co-ethnic politicians (first row) are more positive than mean evaluations of non co-

**The Experimental Stimulus:
Writing a Typical Political Speech**

ethnics (second row). On the other hand, we also expect joking cousins to be evaluated more positively than non-joking cousins, so that mean evaluations of subjects in the first column are more positive than evaluations in the second column. In particular, we expect cousins from a different ethnic group (bottom-left cell) to be evaluated more positively than non-cousins from a different ethnic group (bottom-right cell).¹⁰ Such a finding would be consistent with the idea that due to cousinage relations, members of the same ethnic group have diverse allegiances along a cross-cutting dimension of identity.¹¹

**Table 1: Experimental Design
(Four of Six Treatments)**

	Joking cousins	Not joking cousins
Same Ethnicity		
Different Ethnicity		

We began rolling out this experiment at the end of July 2008; though we have finished initial field-testing at the time of writing, we have not yet seen data from the main phase of data collection. The publication of hypotheses in this newsletter constitutes a public posting of the experimental protocol prior to analysis of the data. Our principal form of analysis for testing these hypotheses will be difference-of-means tests across subjects randomly assigned to each of the six treatment conditions, with ancillary testing of sub-groups due to our interest in possible treatment effect heterogeneity.

In the interest of brevity, I will now describe just two areas in which qualitative methods have been crucial in designing and implementing this experiment: the design of the experimental stimulus, and the creation of a cousinage matrix that allowed us to assign subjects to treatment conditions.

Our goal in designing the experimental stimulus was to create a speech that would engage subjects' attention while mimicking as closely as possible a typical political speech given by a candidate for deputy in the legislature. Here, one of us (Lauren Harrison) drew on earlier fieldwork in which she observed parliamentary campaigns in Bamako in 2007. After comparing our speech to transcripts of real political speeches, we vetted the speech with several Malian informants. I will not belabor the point here but will simply point out that fieldwork and other qualitative methods played an important role in the design of the experimental treatment.

Random Assignment: Creating a Cousinage Matrix

More involved fieldwork was required for the second topic I will discuss here. In order to assign subjects at random to one of the six treatment conditions, we created a large matrix, each row of which corresponds to a Malian last name that we could expect to encounter in the field.

For instance, Table 2 shows a row of the matrix for a person named Keita from the Malinké/Maninka ethnic group. The columns of this row give the last names associated with each of our six treatment conditions. For example, the names in the first two columns are all from the same ethnic group, but Sissoko and Konaté (first column) are considered cousins of the Keita, while Diané (second column) is not. The names in the third and fourth columns, on the other hand, are names associated with other ethnic groups, some of them cousins of the Keita (third column) and some of them not (fourth column). Note that in cells with multiple entries, such as in the first, third, and fourth column in Table 2, the politician's assigned last name was selected at random from the names in the cell.

Table 2: A Typical Row of our Random Assignment Matrix

	(1) Co-ethnic/ Cousin	(2) Co-ethnic/ Not cousin	(3) Not co-ethnic/ Cousin	(4) Not co-ethnic/ Not cousin	(5) No Name	(6) Same Name
Keita (Maninka)	1. Sissoko 2. Konaté	1. Diané	1. Doucouré 2. Sacko 3. Sylla 4. Coulibaly 5. Touré	1. Diallo 2. Cissé 3. Dambélé 4. Théra 5. Dabo 6. Togola 7. Watarra	Pas de nom	Keita (Maninka)

Qualitative fieldwork was crucial for constructing this cousinage matrix. Before arriving in Bamako, we reviewed the secondary literature and conducted interviews with experts on cousinage as well as ordinary Malian informants. This enabled us to determine, as an initial matter, the cousins that are associated with many Malian last names and to construct a preliminary, skeletal matrix. Upon arrival in Mali, we solicited feedback on the matrix from key informants and, with their help, added to the list of names included in the left column (that is, the names of potential subjects) and also refined the list of politicians' names included in each column of each row.

Next, we field-tested an initial version of the matrix on 169 subjects. Data from this initial field trial, as well as additional qualitative information obtained in the field, allowed us to expand and improve the matrix again, and 47 more subjects participated in a second phase of the experiment using our improved matrix. Finally, in mid-August 2008, we revised the matrix once again, for reasons discussed below; this final revised matrix is being used to roll out the experiment during September 2008. Our final version of the matrix includes more than 200 names in the left-hand column, including all of the most typical Malian names.

In our initial field trials, experimental subjects did not always perceive themselves to be in the correct cell—that is, the treatment condition to which they had been randomly assigned. In fact, subjects inferred ethnicity with great accuracy: given only the last name of the politician, and choosing from more than 14 possible ethnic categories, subjects correctly classified the politician's ethnicity 75% of the time. However, in initial trials, they more frequently labeled cousins as non-cousins, or non-cousins as cousins.

This mismatch in initial trials between the treatment conditions to which some subjects were assigned and the treatment conditions they perceived, raises important inferential issues.¹² After all, what we care about in this study is the effect of subject perceptions—we want to know how *perceiving* oneself as being a cousin or not being a cousin of the politician, or his co-ethnic or not, shapes evaluations of the candidate's speech. Here, the mismatch probably occurred for two reasons. First, correctly classifying cousinage relations for over 200 last names is difficult; our initial matrix of cousinage relations was highly imperfect. In this experiment, there was a tradeoff involved in limiting the names of potential subjects. On the one hand, cousinage relations are much better understood by us (and by Malians) for a few very common names, such as Keita, Coulibaly, Touré, or Cissé, than for less common names, so we might have had a better overall accuracy/compliance rate had we limited the study population to subjects with such last names. On the other hand, limiting the number of names would have meant more inefficient and costly subject recruitment.

Second, however, even if we could create a perfectly accurate matrix of cousinage relations, as understood by key informants, people vary in their knowledge of cousinage relations in Mali. For instance, are the Keita and the Doucouré (third column of Table 2) really cousins? Reasonable minds can apparently disagree. As one leading expert on cousinage

puts it, "The question of which *jamu* [patronym] actually jokes with whom is subject to considerable indeterminacy. Lists of the joking partners of any given *jamu* may vary from community to community, or even from individual speaker to speaker" (Launay 2006: 799). Our own experience in the field validated this observation.

The key to resolving this conundrum is that some cousinage links are in fact widely understood: everyone agrees that the Keita and the Coulibaly are cousins. We therefore took the approach of limiting names in the first and third column of Table 2 to those *vrai cousins* or true *senanku* (the Bambara word for cousin), while also only including names in the second and fourth cell that we thought would maximize the chance of correct identification as non-cousins. We devoted considerable effort in the field to accomplishing this task, with the help of key informants. Initial indications suggest that our revised cousinage matrix is allowing much greater accuracy in subject assignment to treatment during the main roll-out of the experiment.

The point is that eliciting a reliable map of cousinage relations from key informants very centrally involved qualitative as well as mixed methods. For instance, to revise our cousinage matrix we conducted qualitative interviews with key informants. We then also employed quantitative analysis of the experimental data from initial trials. To improve the cousinage matrix, we therefore iterated between focused interviews, new versions of the cousinage matrix, and our experimental data to improve the random assignment mechanism in this experiment.

Finally, qualitative methods will likely play a key role in interpreting the results of the experiment—for example, in assessing the extent to which the experimental results can allow us to infer that cousinage plays the political role attributed to it. Here, we will want to analyze the potential role of cousinage in important parliamentary and presidential electoral campaigns.

Conclusion

Natural and field experiments are assuming a place of greater prominence in political science. They also appear to offer substantial opportunities to qualitative researchers. The type of experiment I described in Mali can be implemented relatively inexpensively; in fact, such a project would probably be well within reach for a graduate student working on his or her dissertation. Most importantly for present purposes, natural and field experiments often require skills and case-based knowledge associated with qualitative research. The inferential advantages of natural and field experiments may be increasingly combined with the strengths of qualitative research to generate new forms of mixed-method research, in the service of research programs in many different substantive areas.

Notes

¹ For evidence on the growing use of field and natural experiments, see Gerber and Green (2008) or Dunning (2008).

² Some of the material in this section is based on Dunning (2008);

I am grateful to *Political Research Quarterly* and to co-editor Amy Mazur for permission to use the material.

³ Of course, problems of post-intervention bias can arise: subjects who get the vaccine may tend to go swimming.

⁴ For example, the Keita are part of the Malinké ethnic group, while their joking cousins the Coulibaly are part of the Bambara ethnic group.

⁵ Though Bambara is the first language of one ethnic group in Mali, its use does not imply a particular ethnic identity on the part of the politician. When experimental subjects were not provided with the politician's last name, their guesses about his ethnicity closely tracked the distribution of ethnic groups in Bamako.

⁶ The experimental population is a convenience sample, but distributions on several measured variables are similar to those given by the census for Bamako. However, the experiment under-represents women.

⁷ First name and other identifying information of subjects were subsequently discarded, as described in our protocol approved by Yale's human subjects review board.

⁸ Only experimental subjects could hear the speech through the headphones, and only one subject was recruited from any group; subjects also answered follow-up questions on their own. This limited the potential that subjects' responses to treatment depended on the treatment assignment of other subjects.

⁹ The sixth treatment may allow us to distinguish a "same ethnicity" or a "joking cousin" effect from a mere "sameness" effect: perhaps people simply want to vote for politicians who share their last names.

¹⁰ However, based on our qualitative research, we believe that subjects may not clearly distinguish between cousins and non-cousins, among their co-ethnics.

¹¹ We do not have strong expectations about the sign of any interaction between co-ethnicity and cousinage.

¹² From an experimental design perspective, this issue can be analogized to the problem of compliance with an experimental protocol. See Freedman (2006) for a discussion of relevant analytic approaches.

References

- Brady, Henry E. 2002. "Models for Causal Inference: Going Beyond the Neyman-Rubin Holland Theory." Presented at the Annual Meeting of the APSA Political Methodology Working Group, Seattle, Washington, July 16.
- Brady, Henry E. and John McNulty. 2004. "The Costs of Voting: Evidence from a Natural Experiment." Presented at the Annual Meeting of the Society for Political Methodology, Stanford University, July 29–31, 2004.
- Canut, Cécile, and Étienne Smith. 2006. "Pactes, Alliances et Plaisanteries. Pratiques Locales, Discours Global." In Cécile Canut and Étienne Smith, eds. "Parentés, Plaisanteries et Politique," special issue of *Cahiers D'Études Africaines* XLVI:4, 795–808.
- 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*. Lanham, MD: Rowman & Littlefield.
- Davidheiser, Mark. 2006. "Joking for Peace: Social Organization, Tradition, and Change in Gambian Conflict Management." In Cécile Canut and Étienne Smith, eds. "Parentés, plaisanteries et politique," special issue of *Cahiers D'Études Africaines* XLVI:4, 795–808.
- Dunning, Thad. 2008. "Improving Causal Inference: Strengths and Limitations of Natural Experiments." *Political Research Quarterly* 61:2, 282–93.
- Freedman, David. 1991. "Statistical Models and Shoe Leather." In P.V. Marsden, ed., *Sociological Methodology* 21. Washington, DC: American Sociological Association.

- Freedman, David. 1999. "From Association to Causation: Some Remarks on the History of Statistics." *Statistical Science* 14: 243–58.
- Freedman, David. 2005. *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press.
- Freedman, David. 2006. "Statistical Models for Causation: What Inferential Leverage Do They Provide?" *Evaluation Review* 30: 691–713.
- Freedman, David, Robert Pisani, and Roger Purves. 1997. *Statistics*. 3rd Edition. New York: W.W. Norton, Inc.
- Gerber, Alan S. and Donald P. Green. 2008. "Field Experiments and Natural Experiments." In *The Oxford Handbook of Political Methodology*. Janet Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (New York: Oxford University Press), 357–81.
- Goldthorpe, John. 2001. "Causation, Statistics, and Sociology." *European Sociological Review* 17:1, 1–20.
- Launay, Robert. 2006. "Practical Joking." In Cécile Canut and Étienne Smith, eds. "Parentés, Plaisanteries et Politique," special issue of *Cahiers D'Études Africaines* XLVI:4, 795–808.
- Mauss, Marcel. 1928. "Parentés à Plaisanterie." *Annuaire de l'École pratique des hautes études, section des sciences religieuses* ("Les classiques en sciences sociales") Melun, Imprimerie administrative, Paris: 3–21.
- Posner, Daniel N. 2004. "The Political Salience of Cultural Difference: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi." *American Political Science Review* 98:4, 529–45.
- Radcliffe-Brown, A. R. 1940. "On Joking Relationships." *Africa* 19: 133–40.
- Snow, John. 1855. *On the Mode of Communication of Cholera*. London: Churchill. Reprinted in *Snow on Cholera*, London (Humphrey Milford): Oxford University Press, 1936.

The Promising Integration of Qualitative Methods and Field Experiments

Elizabeth Levy Paluck

Weatherhead Center for International Affairs,
Harvard University
epaluck@wcfia.harvard.edu

Over the past few decades, a productive exchange in political science has explored the idea that qualitative research should be guided by the logic of mainstream quantitative and experimental methods (e.g., Brady and Collier 2004; Gerring and McDermott 2007; King, Keohane, and Verba 1994). Most of these discussions focus on the logic of regression for drawing inferences from observational data, setting aside experimentation as an ideal but rare path to causal inference. A perhaps unintended message of this discussion seems to be that experimentation is a method unrealistic for most qualitative research projects, and consequently, that experimentation is more naturally a quantitative enterprise. In short, qualitative researchers can aspire to use experimental logic for constructing counterfactuals and drawing causal inferences, but cannot use actual experiments.

This essay contends that experimentation, specifically field experimentation, can and should be more central to qualitative research approaches. The argument rests on claims about what

field experimentation is as well as what it is not. Field experimentation *is* one of the strongest methods for inferring causal relationships in real world setting. Field experimentation is *not* inherently quantitative.

By randomly assigning units (individuals, communities, organizations) into two groups, field experiments can infer that differences between the groups are due to an intervening “treatment” (a media program, a land redistribution policy, elite negotiation meetings) applied to one group and not to the other. The key advantage of field experiments is that they draw causal inferences without invoking untestable assumptions that plague observational research about the groups’ *ex ante* comparability.

The most straightforward reason why field experiments are perceived as quantitative enterprises may be found in the psychological concept of the availability heuristic (Kahneman, Slovic, and Tversky 1982). Put simply, there are few available exemplars of field experiments that incorporate qualitative methods or test questions traditionally associated with qualitative investigations, so experiments are thought of as quantitative in nature. I argue that the lack of qualitatively oriented field experiments stems more from our inability to think outside of this heuristic than from unassailable methodological and epistemological divides.

Consider the potential of qualitatively oriented field experimentation using a recent outstanding set of field experiments on gender and political leadership. Chattopadhyay, Duflo, and colleagues (2004; Beaman et al. 2008) capitalized on a policy experiment in India in which the government randomly reserved one third of village council head positions for women (i.e., in councils randomly chosen for a reservation, only a woman could be elected head). The investigators collected primarily quantitative evidence to test the effect of a women leader on public expenditures and on the gender and political attitudes and political behavior of their constituents.

The authors uncovered hugely consequential results: that in some cases women leaders increase women’s political participation, that women leaders distribute public goods differently according to their own preferences (not in response to female constituents’ complaints), and that exposure to a female leader weakens stereotypes about women’s place in the public sphere, but that only after long term exposure does approval of women’s leadership rise.

The point to take from this example and from the rest of this essay is not that qualitative measurement would have made the experimental results “richer” or more detailed, although that is certainly the case. Using qualitative research methods in this field experiment could have provided a different understanding of the causal effect, identified possible causal mechanisms of change, and framed new interpretive understandings of authority, democracy, and gender within an experimentally assessed instance of social change.

For example, participant observation of these women leaders *outside* of the council settings—for example in their homes, where they visit with other women—could have revealed whether they were influenced by women constituents in these more informal settings. Intensive interviews could have

revealed how shifts in beliefs about women leaders’ efficacy may have occurred—for example, did it take public statements of approval from other male council members, or elders or religious leaders? Was there a tipping point mechanism? Such qualitatively generated insights could have enabled this study to contribute more to general theories of identity, leadership, and political and social change. Moreover, ethnographic work could investigate whether understandings of authority and political legitimacy are reshaped in predictable ways by the first few woman leaders, or to what extent narratives about gender are adjusted to fit with women’s new powerful role. Qualitative methods are uniquely positioned to answer these questions.

Potential problems of integrating qualitative methods with field experimentation become apparent from this brief example. For example, many qualitative methods involve more time investment and fieldwork than quantitative data collection; is the extra time feasible or worthwhile in the context of a field experiment? How could participatory or ethnographic methods measure outcomes and processes in a sufficiently large enough sample for an experimental comparison? More challenging, how can field experiments help to investigate traditionally qualitative or observational research questions about historical patterns, institutions, elites, and rare events?

In the rest of this essay, I address these concerns and expand upon some ideas about the integration of field experimentation and qualitative methods and questions. I first describe the benefits of triangulating qualitative and quantitative measurement within a field experiment, using concrete examples from my own experimental work in Central Africa. I argue that qualitative measurement within a field experiment leads to a better understanding of the causal effect, suggests plausible causal explanations and “[extracts] new ideas at close range” (Collier 1999). I next turn to more intensive tools of qualitative inquiry, such as ethnography and interpretive work. These methods can magnify cases, social processes, and concepts within an experiment, and in some cases provide the primary data for causal inference in what Sherman and Strang (2004) term “experimental ethnography.” I explore concerns about small sample sizes and scarcity of available units for random assignment. Finally, I turn to questions traditionally addressed by qualitative and observational research, including questions about historical or rare events. I propose that field experiments have a role to play in many cases, which would require disaggregating complex theories and using theory to specify a universe of cases for present-day experimental tests.

I address this essay not only to qualitative researchers, as encouragement to consider the use of field experimental methods, but also to current field experimentalists, as inspiration to adopt more qualitative approaches in their research.

Triangulating Quantitative and Qualitative Measurement Within Field Experiments

More than the Sum of their Parts

Collecting numerical, categorical, and ordinal data simplifies comparisons between experimental groups, but research

ers could just as well collect and compare qualitative data from interviews, participant observation, and archives. It is widely recognized that inference is best supported by a triangulation of both types of data. Qualitative data can strengthen, modify, or altogether change the interpretation of quantitative data and describe important contemporaneous conditions of change.

The importance of triangulating quantitative evidence with qualitative evidence holds even for a great strength of field experiments—relative to laboratory experiments, field experiments are best positioned to capture effects on real world behaviors because they are located in the behavioral settings of interest (i.e., voting counties, ethnically diverse villages, credit unions). Field experiments capture behavioral data through observational techniques, but most often from public records, such as voting, public expenditure, and police records. Qualitative methods of investigation can explore what these behaviors mean in the context of the study, the possible social and political dynamics by which the behaviors were produced, ripple effects of the changes, and more.

A field experiment I conducted in eastern Democratic Republic of Congo (DRC) randomly assigned half of the radio antennae in the region to broadcast a talk show, which aired and encouraged discussion about a conflict-reduction soap opera broadcast across the entire region (Paluck 2008a). The research question was: Could the talk show increase more face-to-face discussion about conflict reduction, and would this discussion produce more favorable attitudes toward community conflict reduction techniques recommended by the soap opera? The outcome measurement, conducted with individuals in the randomly assigned regions with and without the talk show, included a close-ended survey instrument and a quantitative and qualitative behavioral measure.

In the behavioral measure, surveyors presented each study participant with a two-kilogram bag of salt at the conclusion of the survey. Surveyors told participants the salt was a thank you gift for participating in the interview, and also that a local NGO had identified a group in their community that was in need. Participants were asked if they would like to donate any portion of their salt to this group. “Which group?” participants would ask; surveyors responded per a prewritten script with, “Is there a group you would feel uncomfortable giving this to?” Nearly every participant responded by citing their “least-liked” group:¹ “Yes, the (Banyamulenge/Rega/FDLR).” To this, surveyors responded “Actually, that is the group for whom the donation drive is intended—would you still like to give?” As participants poured some amount of salt into a bag presented by the surveyor, or tied up their bag in preparation to store it away, they discussed their reasons for giving or not giving, their feelings about the donation, their expectations of the consequences of the donation, and their history with the least-liked group. The surveyors recorded this discussion as best they could by hand.

The strength of this mixed qualitative and quantitative measure was four-fold. I was able to record quantitative measures of whether and how much salt each participant gave (which I measured to the gram at the end of each day of inter-

views), qualitative data on the identity of each respondent’s “least-liked” group, and data on participants’ reasoning, feelings, and expectations about helping or not helping this group. To measure the impact of the radio talk show, I used these data and the survey responses to compare listeners in the talk show broadcast regions to listeners in the non-talk show regions.

First I coded the qualitative discussions about the salt, which were fascinatingly diverse: ranging from discussions about norms of sharing (“Congolese must pass on a gift”), to expressions of empathy and perspective taking (“When I am in great need, I know how much help from a stranger means to me”), to strategic reasons (“If I give them this salt, perhaps they will stop targeting my family”), to expressions of pure outrage (“they have killed family members, made us poor—I would rather die than help them”). This information is important and theoretically informative in its own right; I could also correlate their stated reasons and motivations with previous answers to the survey regarding their economic situation, level of education, experience during the ongoing conflict, and other variables.

These qualitative data also significantly strengthened my interpretation of the experimental effect of the radio talk show. Quantitative survey responses showed a *negative* impact of listening to the talk show—talk show listeners compared to listeners in non-talk show regions were less likely to endorse ideas for conflict reduction vis-à-vis their least liked group, and were more likely to endorse statements such as “violence is sometimes necessary in Congolese politics.” The salt measure showed that talk show listeners were also significantly *less* likely to donate their salt to the needy but disliked group (74% percent of control area listeners donated salt, while 55% of talk show area listeners donated). The qualitative discussions pointed in the same (negative) direction as the quantitative survey information regarding the impact of the talk show. I further discovered that radio listeners in the talk show areas expressed significantly more outrage and grievances against the least liked group in their discussions about the salt (controlling for actual reported human rights abuses). The fact that these qualitative data were collected with a different instrument than the quantitative data strengthens support for the inference that encouraging discussion through a radio talk show had a negative impact on listeners. Even stronger triangulation would have included qualitative observations or interviews at another time or in another setting.

Causal Explanation Generation

These qualitative findings suggest a causal explanation for this negative result, specifically that talk show inspired discussion that made grievances more salient to listeners, reminding them of the hurtful actions of the other side. In general, field experimental results become exponentially more useful with these kinds of potential explanations for the process or mechanisms of change. Theory can direct a researcher’s eye toward particular situations and data sources that may explain the causal chain of events, but for more exploratory research (e.g., the effect of childhood abduction into a

militia on voting and political participation as an adult; Blattman 2008), deep contextual absorption (“soaking and poking” in qualitative lexicon) can inductively suggest explanations for experimentally assessed effects.

In the example of child soldiering, Blattman uses semi-structured interviews with former abducted child soldiers, leaders, and social workers to explore explanations for the finding that former child soldiers are more likely to vote. The qualitative data suggest that experience in the militia endowed former child soldiers with a sense of leadership and with a higher degree of maturity (19–20). Causal explanations suggested by such qualitative research can then be tested in successive field experiments. In my own research, I conducted the field experiment in eastern DRC because qualitative research in a previous field experiment testing conflict reduction radio soap operas suggested that discussion was an important mechanism of the observed changes in social norms and behaviors (Paluck, forthcoming). In this previous experiment, I collected systematic observations of groups listening to the treatment and comparison radio programs, and found that listeners kept up a steady rate of interjections, commentary, and side conversations during the broadcast, regarding plot developments and characters’ behavior. Moreover, listeners lingered after the broadcast was over, to share their reactions and digest the messages of the show with one another. I hypothesized that face-to-face discussion about media with community members would shape perceptions of socially acceptable behavior, at least in the confines of that group. The experiment in the DRC attempted to test this causal explanation with an experiment that randomly assigned encouragement to discuss a media program via a talk show.²

In sum, the appeal of field experiments for qualitative researchers is that they offer the opportunity to generate strong causal inferences while “extracting new ideas at close range” (Collier 1999). I suspect that researchers who have long embraced the idea of mixed methodology will readily acknowledge this point. However, despite general enthusiasm for this idea, mixed methods have not been a common feature of field experiments.

Ethnography, Participant Observation, and Interpretive Work

More challenging than combining qualitative and quantitative data within a field experiment is integrating into an experiment qualitative methods that require intensive time investment and field engagement, such as participant observation, intensive interviews, thick description, or ethnography. This broad group of methods is often employed in the service of interpretive goals, for example complicating, historicizing, and enriching understandings of social science concepts like culture, democracy, or power (Wedeen 2002). In some cases, it may be useful to frame interpretive work within a larger field experimental test of an overarching claim of that project. Below I describe how these methods can also be used in a field experiment to investigate causal claims.

One straightforward way to integrate deep interviews, case studies, or ethnographies into an experiment is to select

a reasonable number of observation units for close examination in the experimental and the control groups (see Tarrow 2004, on framing qualitative investigation within quantitative projects). Policy experiments have used this strategy—for example, the Moving to Opportunity experiment in American cities, which tested the effect of giving housing vouchers to low income residents so that they could move into better neighborhoods (Turney et al. 2006). Sociologists and anthropologists working on this project conducted repeated intensive interviews with selected men and women who were randomly assigned to receive or to wait for the voucher. The interviews explored quantitatively measured outcomes such as basic daily functioning and depression, phenomena that often require a fuller contextual understanding. In general, a feasible number of cases for intensive qualitative measurement within an experiment could be selected from each experimental group on the basis of theoretically relevant a priori characteristics to explore the contextual nature and heterogeneity of the experimentally assessed causal effects.

“Experimental Ethnography”

A more ambitious proposal in this vein is to conduct ethnographic case studies for all of the units of observation in a field experiment, or what Sherman and Stang (2004) term “experimental ethnography”:

Experimental ethnography is a tool for answering questions about why programmatic attempts to solve human problems produce what effects, on average, in the context of the strong internal validity of large-sample, randomized, controlled field experiments... This strategy can achieve experiments that create both a strong “black box” test of cause and effect and a rich distillation of how those effects happened inside that black box, person by person, case by case, and story by story (205).

Writing from the perspective of program evaluators, Sherman and Stang discuss a recent randomly implemented policy for restorative justice in England and Australia, which invited victims, perpetrators, and all those affected by the crime to meet and discuss how the perpetrator should repay his or her debt to society. When police officers offered this program to untried perpetrators and their victims, they told each party that if both parties accepted, they would have a 50% chance of having the meeting because the program was in an experimental trial.

Sherman and Stang describe how ethnographies describing the experiences of victims, perpetrators, and their families during and after the restorative justice process would have been important for fully understanding the effects of this program.³ Specifically, experimental ethnography could use an iterative process of theory development and testing commonly associated with qualitative approaches, or grounded theory (Glaser and Straus 1967): “[t]he hypotheses that are generated from interviews or observations of one case can immediately be tested against new data on the same hypotheses collected on other cases. Even if these hypotheses and their tests are later reduced to quantitative form, the fact that they

would not have emerged without ethnographic work provides a strong justification for the added cost and effort of experimental ethnography” (211).

Qualitative data on the severity of the victim’s reaction to the crime, in their example, suggested the hypothesis that the magnitude of potential benefit of restorative justice on the victim’s mental health was directly proportionate to the magnitude of the harm the victim suffered from the crime. The qualitative evidence both “discovered” this grounded claim and offered a way to test it, through continuous comparisons between treatment and control groups. Sherman and Stang note that it is best to conduct this kind of theory testing when *all* of the cases in an experiment can be included in an ethnography, which should be feasible for “samples of a hundred or so” (211).

Small-N Concerns

The Sherman and Stang proposal exposes an important tradeoff, the classic tug of war between breadth and depth that typically leaves qualitative researchers with a small sample size. Other times, qualitative researchers are restricted to a small sample size because of the limited number of units to study—e.g., only six countries that meet the criteria for a certain research topic, or 12 non-overlapping broadcast areas in the region of interest. I have two suggestions regarding this tradeoff.

Collaboration is one answer to the problem of conducting ethnography with all of the units of an experiment. Several qualitative researchers working as a team could each take responsibility for a random sample of units in the treatment and control groups. Researchers’ responsibilities should overlap for a few units, as the overlapping ethnographies could serve as a continuous check on the comparability of their methods and observations. This kind of collaborative ethnography has the potential to provoke a productive discussion among ethnographers regarding the comparative versus particularistic nature of their work. The challenge of comparing their ethnographic data in the service of drawing causal inferences would require ethnographers (or interpreters, participant observation researchers, etc.) to make their process—their definitional terms, their observational procedures, their selection of place and subjects—transparent and replicable. Such an effort would only succeed by increasing the comparative nature of the ethnographic enterprise. While some ethnographic traditions (particularly in anthropology) are opposed to the idea of producing replicable procedures and observations, this kind of a collaborative work would advance the comparative goals of researchers who are amenable to the idea.

A sustained research program is another way to accommodate a small sample in a field experiment.⁴ With a small sample size, researchers may not be able to identify modest or small effects, or may over- or underestimate larger effects. My collaborator Donald Green and I have argued that in this case it is still worthwhile to do the experiment in the context of a sustained research program (Paluck and Green 2008). Repeated experiments on the same general question will average out to the true unbiased effect.

Treating Questions Typically Associated with Observational and Qualitative Investigation

One of the most frequently voiced reasons for not using field experimental methods is that a certain class of research topics are too historically based or would be unethical or impossible to test using random assignment. Questions about the historical pattern of state formation, the causes of revolutions or genocides, elite decision-making about nuclear deterrence, and the democratic peace hypothesis all fall into this category. These topics are sometimes cited as evidence that observational and qualitative researchers struggle with more “important” or “bigger” questions than those addressed by experimental methods.

Of course, field experiments (and as-if-random “natural” experiments; Dunning 2008) have already addressed many important questions that seemed unsuited to experimentation prior to their successful execution. To date, and mostly without the explicit use of qualitative methods, experiments have answered questions about the effects of political campaigns (Green & Gerber 2008; Nickerson 2008; Wanketchon 2003), police raids and crime deterrence (Sherman et al. 2002), mass media programming (Paluck forthcoming; Green and Vavreck 2008), ethnic diversity (Habyarimana et al. 2007; Posner 2004), international election monitoring (Hyde 2007), deliberative democracy techniques (Fararr et al., forthcoming; Wanketchon 2008), gender and politics (Beaman et al. 2008; J. Green 2008), corruption (Olken 2007), employment discrimination (Pager 2007), educational attainment (Sondheimer and Green 2008), health care (King et al. 2007), slavery and trust (Nunn and Wanketchon 2008), and child soldiering (Blattman and Annan 2007). Thus far, I have argued that including qualitative methods can extend the reach of field experimentalism further.

Still, causal questions rooted in history or addressing elites, violence, country-level and rare events like social movements and revolution are at one level beyond the reach of experiments. Random assignment of the purported causes of these events would be unethical or logistically impossible without dictatorial powers or a time machine. One point made in response to this dilemma is that relatively more narrow field experiments accumulate the “stubborn facts that inspire theoretical innovation” (Green 2005). Field experiments gradually collect unbiased causal facts upon which a more complex theory can be built.

I propose another idea that flows in the opposite direction. In contrast to building theories from relatively narrow empirical facts, investigators could start at the level of their highly complex theories and disaggregate them in a way that would make field experimentation possible for a few of the causal links in their specified chain. Theories of genocide, for example, make many causal claims about the road to violence. Some purported causes of genocide include elites threatened by a shift in power, bureaucratic or other tools for ethnic differentiation, land shortages, and so forth. A field experiment could not and would not randomly assign all of these conditions, but it could, for example, examine the effects of policies (introduced progressively in randomly assigned areas of the

country or subsets of the population, i.e., a “random roll out”) that increase or decrease ethnic differentiation (identity cards or citizenship papers), or redistribute land.

Integrating field experiments into these traditionally observational research programs in this manner would require theoretical specificity, strategic case selection (for which qualitative researchers are exceptionally qualified), and (in some cases) cooperation with policy makers or political elites. Researchers would need a high degree of theoretical specificity and clarity in that they would need to define the necessary contextual conditions of a present-day theoretical test. Some theories are intended only for historical cases (Skocpol 1977); in these instances, field experimentation would obviously reach a dead end. But for theories intended to extend into present-day contexts, researchers would need to draw out sufficient and necessary conditions for the field experimental context.

Using theories that describe necessary and sufficient conditions of the phenomenon of interest, qualitative researchers have honed the skill of case selection (Seawright and Gerring 2008) into a systematic method that requires deep contextual and historical knowledge. Selecting present-day relevant cases would be the critical task for researchers for testing theories of historical events with field experiments. Finally, many such field experimental tests would probably require collaboration with policymakers and political elites, since many of these kinds of questions involve structural, economic, or institutional shifts. Many relevant changes are occurring through new policies (again, land policies in developing countries is one example), which could be rolled out randomly. Collaborating with governments and non-governmental or international organizations presents a host of ethical and practical dilemmas, but it should not be written off as impossible. Currently, field experimentation is receiving increased respect and interest from policy makers and international organizations, mostly on the wings of the influential movement to include field experiments in development economic policy and from efforts of some political scientists as well.⁵ As economists have proved with the development community, a few very useful experiments can interest stakeholders in fielding and participating in experiments of their own. Experimentation with (and even on) political elites would make the use of experimentation in observational research programs more of a possibility.

Conclusion

While I do not claim to have all of answers for how qualitative researchers can use experimental methods, I believe that researchers should not foreclose on the possibility of using field experiments in qualitative or observational research programs before considering these ideas. Integrating field experimentation into any research program will be a difficult but creative and productive process. It will require knowledge of the cases, theoretical clarity, and comparable and meaningful outcome measures. Qualitative methods, from case selection to interviewing to participatory observation, are all necessary on some level to conduct good field experiments. For this reason, qualitative researchers and current field experimentalists alike could benefit from considering the ideas I

have reviewed above. Integrating qualitative methods with field experiments should encourage new and interesting investigator collaborations, and learning within all types of methodological persuasions.

Notes

¹ Based on the “content controlled” technique pioneered by Sullivan, Pierson and Marcus (1982).

² Note that in the DRC experiment, I was missing a critical arm of the experiment (due to logistical reasons) in which a third control group did not have access to the soap opera, which provided the topics of discussion, or to the talk show, which encouraged the discussion. Including a no-soap, no-talk control group would show (a) the effect of the soap opera, and (b) the additional effect of discussion inspired by the talk show about the soap opera. I am implementing this design in a new experiment on a peace and democracy radio campaign in Southern Sudan, by randomly assigning a radio show, discussion, radio show plus discussion, or no intervention (Paluck 2008b).

³ They also suggest that ethnographies of the victims and perpetrators who did not accept the offer to be a part of the program would have helped explore the reach of the restorative justice program, and also more generally the ability of experimental trials to measure causal effects in a representative portion of the population.

⁴ Besides the problem of low power to detect causal relationships, small samples mean that simple random assignment is more likely to create an unbalanced comparison. For example, in a sample of twelve manufacturing companies, a random “run” of similar assignment numbers could assign all five car companies in the sample to the treatment condition. This problem of balance can be addressed by matching procedures prior to randomization—simple stratification procedures in which randomization is conducted within stratified groups of car and drug manufacturing companies, for example, or more complex matching with multiple strata using statistical software (e.g., Coarse Exact Matching, Iacus et al. 2008). In my small-*N* experiments in Central Africa, I have randomized within stratified villages and broadcasting regions.

⁵ EGAP, or Experiments on Governance And Politics, is one example of a recent organizational effort involving political scientists and policy organizations.

References

- Beaman, Lori, Raghavendra Chattopadhyay, Rohini Pande and Petia Topalova. 2008. “Powerful Women: Does Exposure Reduce Bias?” Working paper.
- Blattman, Christopher. 2008. “From Violence to Voting: War and Political Participation in Uganda.” Working Paper Number 138, Center for Global Development.
- Blattman, Christopher and Jeannie Annan. 2007. “The Consequences of Child Soldiering.” HiCN Working Paper 22.
- Brady, Henry and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield.
- Chattopadhyay, Raghavendra and Esther Duflo. 2004. “Women as Policy Makers: Evidence from a Randomized Policy Experiment in India.” *Econometrica* 72:5, 1409–43.
- Collier, David. 1999. “Data, Field Work and Extracting New Ideas at Close Range.” *APSA-CP: Newsletter of the APSA Organized Section in Comparative Politics* 10:1 (Winter), 1–2, 4–6.
- Dunning, Thad. 2008. “Improving Causal Inference: Strengths and Limitations of Natural Experiments.” *Political Research Quarterly* 61:2, 282–93.

- Farrar, Cynthia, James Fishkin, Donald P. Green, Christian List, Robert Luskin, and Elizabeth L. Paluck. In press. "Disaggregating Deliberation's Effects: An Experiment Within a Deliberative Poll." *British Journal of Political Science*.
- Gerring, John and Rose McDermott. 2007. "Experiments and Observations: Towards a Unified Framework of Research Design." *American Journal of Political Science* 51 (July), 688–701.
- Glaser, Barney and Anselm Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
- Green, Donald P. 2005. "On Evidence-Based Political Science." *Daedalus* (Summer): 96–100.
- Green, Donald P. and Lynn Vavreck. 2006. "Assessing the Turnout Effects of Rock the Vote's 2004 Television Commercials: A Randomized Field Experiment." Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 20–23, 2006.
- Green, Donald P. and Alan S. Gerber. 2008. *Get Out The Vote: How to Increase Voter Turnout*. Second Edition. Washington: Brookings Institution Press.
- Green, Jennifer. 2008. "Mobilizing Women to Vote in Traditional Societies: An Experiment Encouraging Political Participation in Rural India." Working paper.
- Habyarimana, James, Macartan Humphreys, Daniel Posner, and Jeremy Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101:4, 709–25.
- Hyde, Susan. 2007. "The Observer Effect in International Politics: Evidence from a Natural Experiment." *World Politics* 60:1, 37–63.
- Iacus, Stefano M., Gary King, and Giuseppe Porro, "Matching for Causal Inference Without Balance Checking." Harvard University Working Paper, available at <http://gking.harvard.edu/files/abs/cem-abs.shtml>.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.
- King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T. Moore, Jason Lakin, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila, and Héctor Hernández Llamas. "A 'Politically Robust' Experimental Design for Public Policy Evaluation, with Application to the Mexican Universal Health Insurance Program." *Journal of Policy Analysis and Management* 26:3, 479–506.
- Nickerson, David W. 2008. "Is Voting Contagious? Evidence from Two Field Experiments." *American Political Science Review* 102 (February), 49–57.
- Nunn, Nathan and Leonard Wantchekon. 2008. "The Trans-Atlantic Slave Trade and the Historical Origins of Mistrust in Africa: An Empirical Analysis." Working paper.
- Olken, Benjamin. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (April), 200–49.
- Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." *Annals of the American Academy of Political and Social Science* 609, 104–33.
- Paluck, Elizabeth Levy. Forthcoming. "Reducing Intergroup Prejudice and Conflict Using the Media: A Field Experiment in Rwanda." *Journal of Personality and Social Psychology*.
- Paluck, Elizabeth Levy. 2008a. "Is it Better Not to Talk? A Field Experiment on Talk Radio and Ethnic Relations in Eastern Democratic Republic of Congo." Working paper, Harvard University.
- Paluck, Elizabeth Levy. 2008b. "A Field Experiment Testing Discussion and Media in Southern Sudan." Unpublished Paper.
- Paluck, Elizabeth Levy and Donald P. Green. 2008. "Deference, Dissent, and Dispute Resolution: A Field Experiment on a Mass Media Intervention in Rwanda." Unpublished Paper.
- Posner, Daniel. 2004. "The Political Salience of Cultural Difference: Why Chewas and Tumbukas are Allies in Zambia and Adversaries in Malawi." *American Political Science Review* 98:4, 529–45.
- Seawright, Jason and John Gerring. 2008. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61:2, 294–308.
- Sherman, Lawrence and Heather Strang. 2004. "Experimental Ethnography: The Marriage of Qualitative And Quantitative Research." *The Annals of the American Academy of Political and Social Sciences* 595, 204–22.
- Sherman, Lawrence, Dennis Rogan, Timothy Edwards, Rachel Whipple, Dennis Schreve, Daniel Witcher, William Trimble, Robert Velke, Mark Blumberg, Anne Beatty, and Carol Bridgeforth. 2002. "Deterrent Effects of Police Raids on Crack Houses: A Randomized, Controlled Experiment." *Justice Quarterly* 12:4, 755–81.
- Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.
- Sondheimer, Rachel M. and Donald Green. 2008. "The Brody Paradox Revisited: Using Experiments to Estimate the Effects of Education on Voter Turnout." Unpublished paper.
- Sullivan, John L., James Piereson, and George E. Marcus. 1993. *Political Tolerance and American Democracy*. Chicago: University of Chicago Press.
- Tarrow, Sidney. 2004. "Bridging the Quantitative-Qualitative Divide." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Henry E. Brady and David Collier, eds. (Lanham, MD: Rowman & Littlefield), 171–80.
- Turney, Kristin, Susan Clampet-Lundquist, Kathryn Edin, Jeffrey R. Kling, and Greg J. Duncan. 2006. "Neighborhood effects on Barriers to Employment: Results from a Randomized Housing Mobility Experiment in Baltimore." Working paper #511, Princeton University.
- Wantchekon, Leonard. 2003. "Clientalism and Voting Behavior: Evidence from a Field Experiment in Benin." *World Politics* 55 (April), 399–422.
- Wantchekon, Leonard. 2008. "Expert Information, Public Deliberation, and Electoral Support for "Good" Governance: Experimental Evidence from Benin." Working paper.
- Wedeen, Lisa. 2002. "Conceptualizing Culture: Possibilities for Political Science." *American Political Science Review* 96:4 (December), 713–28.

Battling Onward: The Debate Over Field Research in Developmental Economics and its Implications for Comparative Politics

Edmund J. Malesky

University of California, San Diego
emalesky@ucsd.edu

Victory has finally been declared after a fierce internecine struggle, but the conflict will never be recorded in the Correlates of War dataset. In fact, except for a few notable exceptions, news of the battle has hardly seeped into political science journals at all; and its impact has certainly not been appreciated by the group best positioned to capitalize on the terms of victory—comparative scholars with deep country knowledge.

The war has been fought on the pages of economics journals¹ and the main battle lines have been drawn over the standards of admissible evidence used to assess theories of economic development. On one side were practitioners of old-school research, who predominantly studied economic growth, trade, and capital flows, and for whom the major research tool has been large-*N*, cross-national regressions (Rodrik 2008). On the other side were practitioners of the New School of Developmental Economics (NSDE). These researchers, sometimes referred to sardonically as “randomistas,” abhor cross-national designs, distrust high-tech statistical band-aids, and believe the only evidence that is worth paying attention to are randomized field trials, where the impact of treatments can be readily observed, or natural experiments, where accidents of history mimic randomized trials (Banerjee 2007). Although there is still some dissent among those whose research projects do not easily lend themselves to these methods, NSDE tools are now considered to be the cutting edge of economics research; their work dominates major economics publications and they have been major beneficiaries of international and federal research funding.²

The logic of the NSDE research agenda can be applied to a number of research arenas that political scientists are interested in as well. As I will argue below, comparative politics is a particularly target-rich environment. Moreover, NSDE approaches significantly alter the entry costs of empirical work. Mathematical facility is relatively less important, while a premium is applied to detailed country knowledge that allows researchers to identify appropriate natural experiments and use contacts to achieve buy-in on conducting gold-standard policy interventions. In short, comparative political scientists are better-suited than other subfields to begin applying these techniques in our work. In this essay, I introduce the NSDE, discuss their most prominent methods, and assess pros and cons of applying this approach to fieldwork in comparative politics.

Tools of the New Developmental Economists

NSDE congratulates their forefathers in developmental economics for developing a large number of important theories, but they argue that these theories have been tested very poorly, if at all. First, they argue that large-*N*, cross-national studies are insufficient for addressing the detailed micro-logic implied by economists’ formal models. Why focus our research on aggregate associations, when a theory specifically involves individual behavior? For instance, an argument that links property rights development and economic growth implies that individual businessmen are expanding investment because they feel more secure. Thus, a positive exogenous shock to property rights allocation should be followed by greater investment at the firm-level. It is this firm-level behavior that should be the focus of investigation.

Second, NSDE practitioners argue that measurement of key theoretical concepts has traditionally been slapdash and that better fieldwork can offer more precise measurement testing. Ray Fisman’s 2001 study of the impact of political connections and business performance in Suharto’s Indonesia was a landmark example of the dividends of concern for careful operationalization. Fisman demonstrated the importance of political connections by studying the performance of listed companies on days that the media reported news about Suharto’s failing health. He found that companies closely associated with Suharto, his family members, or high-ranking Golkar party members dropped precipitously after such events, when compared to other listings.

Third, they argue that their predecessors have been sloppy about *identification*, by which they mean resolving endogeneity and selection bias in their empirical tests. It is worth dwelling on what Abjhit Banerjee (2005: 31) has termed the “sacrifices to the harsh god of identification,” because it is really the motivating principle of NSDE, the flag around which they all rally. Their ferocity in identification stands in sharp contrast to the cavalier approach with which it has been regarded in comparative politics to date.³ King, Keohane, and Verba (1994), of course, described the problem of endogeneity in their seminal book on methodology in social science, but their main take-home rule was to avoid it in designing a research project. This can be less than helpful to new graduate students, because many of the big theories of comparative politics that fascinate us have some element of endogeneity. We are therefore forced to: (1) ignore the critical questions of the discipline, which rarely makes for a successful dissertation; (2) wave our hands, declare “everything is endogenous,” and push onward with research project that is flawed at its core; (3) lag our key causal variable in regression analysis and claim that this accounts for reverse causality, while praying that reviewers do not call us on the sleight of hand; or (4) use a mixed-methods approach that uses case studies to process trace the true sequencing of causes and outcomes (Gerring 2007: 37). None of the first three options would past muster with NSDE and option four can become easily bogged down in debates about whether the case selection was appropriate for the question, generalizable beyond the specific instance,

and avoids biased historiography in the choice of sources.

For NSDE, adequate identification strategies require one of three techniques, which I list in order of their favor by the field:

*Field experiments, where the researcher treats a randomly selected sub-sample with a particular policy and therefore knows for certain that policy was exogenous.*⁴ An excellent example of this type of work is Ben Olken's 2007 analysis of corruption in Indonesia. A long-held theory in the discipline and among practitioners was that decentralization of responsibilities to subnational governments, combined with participatory oversight mechanisms (e.g., town-hall meetings) limited corruption.

Previous economists and political scientists studied this question, but the primary analysis was to regress a cross-national indicator of corruption on cross-national indicators of levels of administration decentralization (See Rose-Ackerman 2004 for a review of this work). Olken was not satisfied with these results because the cross-national measures of corruption perception (often by foreign investors) were not precise enough to link them directly to the decision to decentralize. Alternatively, organizations like the World Bank (2004) had developed case studies of grassroots accountability mechanisms, but these studies were plagued by selection bias. Sites had been specifically selected due their good governance relative to their peers.

Unsatisfied with previous work, Olken put this theory to the test by convincing the Indonesian office of the World Bank to randomize its oversight over six hundred village-level road-building projects. All villages received a similar pot of money for road construction, but Olken randomized three accountability mechanisms. Some villages were trained in the art of town-hall meetings and local elections to the town council. Other villages were threatened with an audit by a central regulatory agency. And a third group was not provided with any accountability mechanism at all.

Olken then built test roads in order to determine how much the materials cost to build a kilometer of road. Using a team of engineers, Olken then hired a team of engineers to take core samples of each of the 600 village-level road projects. He found that local participation through town councils was not significantly different from no monitoring at all. Threat of a central audit, however, did significantly reduce theft of road materials.

Olken's work undermined a key argument in theories of corruption monitoring. It is a critical finding that continues to reverberate in development circles. What made it possible was Olken's detailed understanding of Indonesia that allowed him to identify the most precise way to measure corruption and to identify organizations that were willing to have their policies evaluated.

Discovery of an ideal and verifiable natural experiment that provides an exogenous shock to the key causal variable. Sometimes, despite a researcher's best persuasive efforts, it is impossible to convince governments and donors to randomize a policy treatment, especially in the areas that concern political scientists. In other cases, career motivations prevent government officials and donors from accepting randomized evalua-

tions of their interventions. In these cases, I have found the new developmental economists to be at their creative best in identifying shocks which function as natural experiments, creating an exogenous and quasi-random treatment that can be exploited (Dunning 2008b).

My favorite example of this is the brilliant work on child soldiering by Blattman and Annan (2006), because they tackle an incredibly complex issue that is of direct interest to comparative political scientists: what are the economic and psychological consequences of child soldiering? Selection bias poses a clear problem for this work; perhaps the same traits that make a child attractive for abduction into child soldiering also impact their chances of post-war success. Or perhaps militias focused their abductions on particular ethnic groups, which were already disadvantaged economically.

To resolve this problem, Blattman and Annan explore the methods of abduction, discovering that they primarily occur through raids on isolated minority villages. As a result, the choice of children is haphazard, depending solely on who could not manage to escape at the time of the raid. To construct their sample, they select abducted and un-abducted siblings from the same household, which allows him to hold constant most demographic features that would confound analysis. Using this creative strategy, the authors find that child soldiers are less likely to finish elementary education, achieve functional literacy, and obtain skill-intensive labor in later life.

Occasionally, scholars dig deep into history to identify an appropriate natural experiment. Banarjee and Iyer (2006), for instance, use British patterns of land administration in colonial India as an exogenous treatment for institutional development. Different British colonial administrators implemented either a landlord based system or a land tenure system in different regions of India. They find that that original treatment leads to vast differences in electoral participation, public goods provision, and ultimately economic growth.

Once again, these findings do not involve mathematically complex approaches; it is the area-studies knowledge of the researchers and their passion for particular issues that allowed them to exploit the opportunities that were presented them.

Discovery of the ideal instrumental variable, which can be employed in a two-stage procedure to resolve endogeneity bias. The worst-case scenario for NSDE is to fall back on an instrumental variable. Here, randomization to address endogeneity and unobserved heterogeneity has essentially been taken out of the scholars hands and a suitable exogenous treatment (that simulates random assignment) cannot be identified. Usually, a scholar resorts to this technique when the question is considered to be of critical importance to economic theory or policy choice. In these cases, the goal of instrumental variable selection is essentially to simulate a natural experiment by identifying the exogenous portion of a key causal variable and testing the impact of that exogenous component on the outcome of interest.

Instrumental variables are common in political science as well, but what differentiates the developmental economists is their explicit consideration of IV-regression as a substitute for a natural experiment and the care they give to interpreting the

results (well...sometimes).

By far, the most famous exemplar of this type of analysis is Angrist's (1990) study of the impact of military service on lifetime earnings. Of course, the choice of a military career is not exogenous, so disentangling the factors that lead someone to choose military over civilian employment is nearly impossible using the data available on social security records. Angrist, however, knew that a great deal of service in Vietnam took place as a result of a draft-era lottery system where soldiers were selected by birth date. Because a lottery is by definition random, Angrist could use birth date as an instrument to identify the effect of military service, finding significant and negative results.

Most applications of instrumental variables would have stopped there, but Angrist created a new benchmark for the field by defining the Local Average Treatment Effect (LATE) of his instrumental variable selection. Because Angrist and the New Developmental economics school see instrumentation as a poor man's substitute for randomized treatment, they explicitly seek to identify the parameters of their imperfect treatment. Angrist recognized that the birth-date lottery was a treatment that did not apply to all the observations in his sample. Specifically, two groups were excluded: soldiers who self-enlisted in military service and draft dodgers who shirked their randomly selected responsibility. Thus, he stated unequivocally that his findings cannot be applied to these sub-samples; they are limited to the general populace who may or may not have received the lottery treatment.

The concept of LATE is often extremely helpful for understanding why point predictions and the impact of control variables can change dramatically in an IV model, even if the instrument selected is strong and meets the exclusion criteria. The LATE allows readers to understand the sub-population to which the instrumented treatment applies.

Another excellent example of the careful NSDE approach to instruments is Woodruff and Zenteno (2007). In this clever piece, the authors use distance from the railroad in Mexico to identify migration and consequently study its impact on micro-enterprises in Mexico. The railroad instrument works because Woodruff and Zenteno are able to show how proximity to a railroad station incentivized the decision to migrate, but had little impact on the choice to start a new business. Once again, the two scholars nail down the LATE, so the reader understands the parameters of the analysis.

For all three above techniques, panel models that study the same observations over time are requisite; single-shot surveys and data collection are uniformly thought of as inappropriate for identification. Only through panels can a researcher truly identify the hypothesized changes in time resulting from the treatment. At the very least, scholars should try to construct retrospective panels from existing surveys, but these are imperfect due to the memory deficiencies of respondents.⁵

Can these Techniques be Applied Successfully in Comparative Politics?

Absolutely; in fact, it is already beginning to take place in a number of interesting arenas. A recent issue of the *American*

Political Science Review (102:1, 2008) featured three field experiments and a survey experiment among its ten articles. In comparative politics, Donald Green, James Gibson, Daniel Posner, Jeremy Weinstein, Marc Cartan Humphries, Leonard Wantchekon, Susan Hyde, and Elizabeth Paluck have been at the forefront of using these NSDE approaches to study explicitly political questions.⁶

But there are still many more opportunities. The same issues which spurred on the new developmental economists are currently present in political science. Just like the economics literature, comparative politics has a range of important theories that are built upon the behavior of individual actors, but have only been tested in aggregate cross-national analyses. As with any new approach, there are pros and cons to adoption. The individual actions of voters, interest groups, and political entrepreneurs buttress a number of the cross-national associations that have been discovered in our discipline. Whether these actors really behave in the manner assumed remains an open question.

The Benefits of a Developmental Economics Approach for Mixed-Method Researchers

The most important benefit of following developmental economists in their approach is the precision of the analysis. Scholars are removed from the well-known difficulties of studying large cross-national datasets, where measurement error, unobserved heterogeneity, and endogeneity frustrate even the most theoretically sound projects. Political science has recently been at the forefront of devising new and more sophisticated approaches to dealing with unsatisfactory data. Our graduate students today must effectively minor in statistics in order to address the myriad problems hidden in datasets such as Polity IV or the World Values Survey. But as Donald Green (2005) has pointed out, observational analyses often perform quite poorly relative to experimental analyses of the same issues, even when observational scholars use cutting-edge statistical specifications.

For NSDE, the costs of analysis are on the front end. If a field experiment is well designed or a natural experiment carefully chosen, a scholar can analyze the resulting data with comparably easy statistical operations. In fact, simpler is generally considered to be better. A sure-fire way to engender suspicion in an NSDE audience is to present an empirical result that shows up after a sophisticated procedure (i.e., error correction models or generalized method of moments (GMM) regression) that does not hold using simpler specifications. Simplicity also has the dual advantage of allowing the research to be accessible to non-quantitative scholars, which should go a long way toward uniting our divided field. Particularly because, as Elizabeth Paluck argues in her contribution to this symposium, experiments generate excellent qualitative as well as quantitative information.

A second benefit is that it plays to the strengths of comparative political scientists. We often know these countries better than anyone else, speak the local language, and possess a range of contacts in government, business, and development sectors. Often, our fascination with particular coun-

tries is what drove us to study comparative political science in the first place. Because of this, we are ideally suited to identify appropriate national experiments or convince practitioners (policy makers or donors) to randomize a policy intervention. It is ironic that we have surrendered this ground to the “number-crunching” economists.

We shouldn't be dismayed that economists got the jump on us, however, because many of their randomizations allow for what Green has termed “downstream experiments,” where the same randomization can allow for follow-on studies (2005). Green points out that educational attainment may have been initially studied by economists—through randomly distributed college scholarships—to study its impact on variables of economic interest, but we can return to this sample later on to study education's effect on election turnout, racial tolerance, civil engagement, or job performance.

The Negative Side of NSDE

Before drinking the NSDE Kool-Aid, it is worthwhile to highlight some limitations of the approach. NSDE scholars actively discuss these issues, of course, but there is still work to do in resolving them.

First, a focus on the techniques may force scholars to avoid big questions. As discussed above, the ideal target for NSDE tools in comparative politics is when a large theory implies specific individual-level behavior. Creative scholars can focus on the micro-logic and identify a test of it. In KKV (1994) terminology, they can test the individual observable implications of the larger theory. If the theory is an explicitly macro theory about the behavior of states in the world system or parties within political systems, cute identification strategies may not be available. Rare is the government that allows a researcher to randomly sow ethnic strife across subnational entities. And domestic terrorist organizations, a hot topic of current scholarship, are too few in number and too strategic in their choice of targets to see if a natural experiment is applicable.

NSDE scholars would certainly retort that my characterization is unfair and that many macro-theories have individual implications if you think carefully about the design. For instance, instead of focusing on states as the unit of analysis, one could focus on specific policies, legislative documents, or exploit subnational variation. Yet even diehards would concede that there are certain issues that do not lend themselves well to gold-standard randomization. Here, the best a scholar can do is hope to identify an instrumental variable for the task at hand. And, well...this can sometimes be like waiting for Prince Charming to arrive. The instrument must meet some very high standards (known as the exclusion criteria), must be suitably strong, and must survive a range of challenging (and exhausting) diagnostic tests.⁷ And that is just for cross-sectional analysis. In a panel model, Prince Charming must vary over time! What is the conscientious scholar to do?

A second problem is the generalizability of the research findings. NSDE scholars unapologetically “go small” in their research designs. Rather than controlling *ex post* for confounding factors, they select them *ex ante* to minimize observed

heterogeneity, hoping their randomization procedures will address unobserved heterogeneity. But how much analytical leverage do we really get from a few hundred villages in Indonesia or India? This criticism is very similar to those leveled against detailed case-study work. How can we be sure that the results of a theoretical test will travel to other developing countries, or even developed countries (where field experiments are rarely attempted)? Quite simply, we don't; and the vagaries of academia mean we probably never will. Clever field experiments and natural experiments get honored only once in a top-shelf publications. Applications of the design to other contexts (normal science in a Kuhnian sense) are simply not rewarded by the field.

NSDE chooses to privilege small studies executed to near perfection with limited external validity over large, observational studies that can never be executed perfectly but have some external validity. As Banarjee (2005) puts it, “...even if we have many low quality regressions that say the same thing, there is no sense in which the high quality evidence becomes irrelevant—after all, the same source of bias could be affecting all the low quality results.” In essence, Banarjee believes the inference problems from poorly designed, large-*N* works are so severe that the external validity they purport to offer should be discounted. Actual learning, however incremental, is preferable to the illusion of learning that observational studies offer (Green 2005).

Third, banking on an NSDE technique early in an academic career can be dangerous, because the choice to use them is not always in scholars' hands. In the case of field experiments, government officials and donors must agree to a randomized policy intervention, which does not always coincide with their own career incentives. It can be much easier to pre-select areas where one knows a controversial policy will work than randomize and remove all doubt. There is no doubt that fewer of my e-mails to Vietnamese and Cambodian practitioners are returned since I began pushing randomized approaches.

Similarly, natural experiments may not occur exactly where a research question demands they do. This can create an odd form of inductive research, where a researcher identifies an exogenous shock and then works backwards to identify a question that can be answered using it. I think we all can agree that is not a practice that should be condoned by our discipline. Scholarship advances with the work of creative puzzle solvers, not puzzle hunters.

Finally, I have one word of warning for scholars who are fortunate enough to be able to use an NSDE technique for their research question. Randomization and exogenous shocks do not excuse sloppiness on other aspects of the research program. The vast majority of experiments rely on some form of survey to measure the dependent variable. In studying some of these survey instruments in preparation for a course on research methods, I was shocked to discover that many of the same issues which contaminate general survey research can be found in these well-designed studies. Double-barreled questions, ambiguous terminology, and potential framing effects in question ordering are rife in many instruments. Biases such as

these, which are often exacerbated by unclear manuals for hired interviewers, can be just as detrimental to inference as the problems NSDE techniques were designed to solve. It is amazing that scholars can spend so much time carefully identifying the perfect randomized intervention but create a messy survey to analyze it.

Conclusion

Despite my warnings, I generally believe NSDE techniques are good for the field. It is time that comparative politics also rallied around the flag of identification. There are a number of juicy targets available to explore with more careful research designs. If we do not study them, the economists certainly will. Economics long ago abandoned the idea that their discipline was devoted to the analysis of money. It is now a discipline of tools in search of questions, and comparative politics has some of the most fascinating queries.

I also feel strongly that all four problems I identified with NSDE can be solved. More macro-theories will be felled once identification becomes mainstream in political science and more creative minds go to work looking for angles to cut into those theories. Generalizability has already been identified as a concern by NSDE, and organizations such as Dean Karlan's Innovations for Poverty and the Jameel Poverty Action Lab are hard at work chronicling randomized experiments and, recognizing that there is little academic benefit to replication, pushing international donors to replicate randomized experiments in new contexts. If big donors can be convinced, there will be many more opportunities for such research. There is already some progress. The Millennium Challenge Corporation and the World Bank Group's Private Sector Development Facilities have NSDE approaches as part of their organizational mandates. Finally, sloppy survey work can and should be rectified easily.

In short, I am positive about bringing NSDE to field research in comparative politics. Now, we just need to take the battle forward and assign more of these articles in our methods classes.

Notes

¹ And I imagine, although I have no definitive proof, in the review letters on submissions to those journals.

² For the official declaration of victory and the discussion of further territories to be conquered see "New Directions in Developmental Economics: Theory or Empirics? A Symposium." *Economic and Political Weekly*. August 2005. <http://www.arts.cornell.edu/poverty/kanbur/NewDirectionsDevEcon.pdf>.

³ In less polite conversations, the NSDE has been referred to as the "Identification Taliban."

⁴ See Duflo, Glennerster, and Kremer (2006) for a handy primer on how to implement such projects.

⁵ NSDE scholars are not the only group concerned about inference. Other disciplines (education and psychology) have developed their own, less field-intensive modes of resolving these problems. These include direct matching, propensity score matching, and most recently, synthetic control methods for comparative case studies (see Abadie and Gardeazabal 2003).

⁶ Why African specialists are at the forefront of this approach is a research question in its own right.

⁷ For a nice discussion of these and other principles, see Dunning (2008a) and Murray (2006).

References

- Abadie, Alberto and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93:1, 112–32.
- Angrist, Joshua D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *American Economic Review* 80:3, 313–36.
- Banerjee, Abhijit and Lakshmi Iyer. 2005. "History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review* 95:4, 1190–1213.
- Banerjee, Abhijit. 2005. "New Development Economics" and the Challenge to Theory." *Economic and Political Weekly* (August): 30–39.
- Blattman, Christopher and Jeannie Annan. 2007. "The Consequences of Child Soldiering." HiCN Working Paper #22. <http://www.hicn.org/papers/wp22.pdf>.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2006. "Using Randomization in Development Economics Research: A Toolkit." CEPR Discussion Papers 6059, December.
- Dunning, Thad. 2008a. "Model Specification in Instrumental Variables Regression." *Political Analysis* 16:3, 290–302.
- Dunning, Thad. 2008b. "Improving Causal Inference: Strengths and Limitations of Natural Experiments." *Political Research Quarterly* 61:2, 282–93.
- Fisman, Raymond. 2001. "Estimating the Value of Political Connections." *American Economic Review* 91:4, 1095–1102.
- Gerring, John. 2007. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Gibson, James. 2008. "Group Identities and Theories of Justice: An Experimental Investigation into the Justice and Injustice of Land Squatting in South Africa." *Journal of Politics* 70:3, 200–16.
- Green, Donald. 2005. "The Illusion of Learning." *Deadalus* (Summer): 97–99.
- Habyarimana, James Macartan Humphreys, and Jeremy Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101:4, 709–25.
- Hyde, Susan. 2005. "Introducing Randomization to International Election Observation: The 2004 Presidential Elections in Indonesia." Dissertation, University of California, San Diego, Chapter 6.
- King, Gary, Robert Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University.
- Rodrik, Dani. 2008. "We Shall Experiment, but How Shall We Learn?" Prepared for the Brookings Development Conference, May 29–30.
- Wantchekon, Leonard. 2003. "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin." *World Politics* 55:3, 399–422.
- Woodruff, Christopher and Rene Zenteno. 2007. "Migration Networks and Microenterprises in Mexico." *Journal of Development Economics* 82:2, 509–28.
- World Bank. 2004. *Making Services Work for Poor People*. World Development Report 2004. Washington, DC: World Bank.

The Contribution of Area Studies

Stephen E. Hanson

University of Washington
shanson@u.washington.edu

The role of “area studies” in the comparative politics subfield has been the subject of a prolonged and acrimonious debate.¹ On one hand, advocates of a more deductive approach to social science inference have railed against area specialists for their presumed hostility to both generalizable theory and quantitative methodology. As Robert Bates warned in his initial “Letter from the President” in the APSA Comparative Politics Newsletter:

Within the academy, the consensus has formed that area studies has failed to generate scientific knowledge. Many see area specialists as having defected from the social sciences to the camp of the humanists. . . . They tend to lag behind others in terms of their knowledge of statistics, their commitment to theory, and their familiarity with mathematical approaches to the study of politics (Bates 1996).

Area studies centers, Bates argues, thus represent something of an institutional “problem for political science,” given their ability to attract sufficient outside funding and administrative support to allow area specialists to operate independently of departmental constraints. Bates does concede the continuing relevance of the work of area specialists, but only insofar as they continue to “record the data from which political inferences [might] be drawn by social scientists residing in political science departments” (ibid.). Some supporters of Bates’s position even suggested that this proposed division of labor could become an international one, since “the area expertise of natives working in their own countries is likely to be deeper and richer than what those of us working in the U.S. can generate anyway,” while more theoretically sophisticated and methodologically rigorous scholars in the U.S. might increasingly focus on “non-specific, increasingly abstract ‘theory’ ” (Golden 1998: 6).

On the other hand, defenders of area studies have insisted that political analysis should always be grounded in a thorough knowledge of regional context, without which, in their view, a deeper understanding of the dynamics of political order and political change is impossible. Attacks on area studies, according to this camp, reflect the ascendance of narrowly economic approaches to politics that reflect a parochial misunderstanding of human motivation. Thus Chalmers Johnson excoriates Bates for assuming that “rational choice theory contains a unique capacity to transcend culture and reduce all human behavior to a few individual motivational uniformities” (Johnson 2001: 60). As for Bates’s and Golden’s idea that area specialists might play a useful role as purveyors of regional data useful for more general theoretical analysis, Johnson replies acidly: “One problem with this proposed division of labor is that these social scientists do not produce beautiful objects but junk and real area specialists have a much better

record of producing theory than their self-proclaimed theoretical rivals” (Johnson 2001: 62).

Given the invidious tenor of this debate, one might think that the battle over the role of area studies in the comparative politics subfield was truly unprecedented. When one steps back to examine the entire history of the interaction between area studies and the political science discipline, however, one quickly discovers that the basic contours of this debate are very old indeed. Note, for example, the characterization of the area studies controversy contained in the 1973 APSA Presidential Address of prominent Japan specialist Robert E. Ward of Stanford University:

Many behavioralists allege that something called the ‘Area Approach’ lacks rigor and scientific potentiality. It is viewed as descriptive and relativistic, often historical or institutional rather than behavioral in focus, and normally idiosyncratic in terms of its findings. In current professional parlance these are not terms of praise.

A few so-called ‘area types’ respond with countercharges of cultural illiteracy, gross ethnocentrism, uncritical scientism, and scornful characterizations of those members of the opposition who venture abroad as ‘itinerant methodologists’ or worse (Ward 1975: 27).

With only a few minor terminological updates—the word “behavioral,” which used to signify scientific rigor, would now be changed to “analytical”; the word “institutional” would now be claimed by the champions of deductive theory—the same speech could be delivered again today, over three decades later, practically verbatim.

This is quite remarkable, given that Ward’s speech emphasized that this fruitless debate had already been raging in the discipline for many decades, and that some way to transcend it should therefore be urgently sought. Indeed, Ward began his presentation by quoting Jasper Shannon’s 1950 presidential address to the Southern Political Science Association, which detailed with ironic humor how, in the history of political science since the era of Woodrow Wilson, initial optimism about the discovery of “laws of politics as clear and definite as those of physics”—knowledge of which would enable the spread of good institutions across the world—had gradually given way in the face of the twin disasters of the Great Depression and World War II to increasing skepticism about the possibility of arriving at a single neutral science of politics (Shannon, quoted in Ward 1975: 24–26).

Thus, with some slight thematic and stylistic variations, the political science debate about the contribution of area studies has raged in much the same form for nearly a century now. It is my contention here, however, that the puzzling longevity of the area studies controversy provides an important clue as to its nature and possible resolution. Specifically, I argue that the area studies debate reflects a deeper confusion among political scientists of all stripes about how to combine deductive and inductive reasoning in social research. To transcend this debate will require scholarly agreement on some difficult and fundamental questions about how to develop scientifically fruitful typologies of regime type—agreement

that is highly unlikely in the absence of mutual intellectual respect between self-professed area specialists and comparative politics generalists.

Stereotypes about Area Studies

Before we can assess the contribution of area studies to the subfield of comparative politics, we need to understand precisely how “area studies” as a distinct approach to the analysis of world politics emerged in the first place. In this respect, several common criticisms of the area studies approach appear to rest on faulty historical premises.

Thus, one common criticism of area studies is that its particular orientations and choice of material were directly influenced by the Cold War priorities of the United States government—in particular, the global struggle against communism (Cohen 1985; Farish 2005). Advocates of this viewpoint point out that the initial push to expand area studies training in the postwar era by the Rockefeller, Ford, and Carnegie Foundations focused primarily on Russia and East Asia—the early sites of conflict between the United States and its communist adversaries. The National Defense Education Act of 1958, which initiated the Title VI National Resource Centers of the Department of Education that provide the basic infrastructure for area studies training in most major U.S. universities to this day, was launched as a direct response to the Soviet launch of the *Sputnik* satellite the year before. Many leading area specialists in political science departments spent time in the U.S. military during World War II, and many more lent their expertise to various U.S. government security agencies throughout the Cold War period (Lambert et al. 1984: 2–10).

The connection between the funding of area studies programs and the strategic priorities of the U.S. government cannot be denied.² However, the further argument that U.S. and private funding of area studies centers undermined their scholarly objectivity assumes that the divide between political science generalists and area studies practitioners overlaps substantially with an ideological divide between critics and defenders of U.S. foreign policy. Closer examination of the issue calls such an assumption into question. From Herman Kahn to Steven Krasner, social scientists explicitly oriented toward general social science theory have been just as likely to be engaged in the formulation of U.S. foreign policy as have prominent area specialists. Among the latter, too, one finds any number of critics of U.S. policy making, ranging from the neo-Marxist Barrington Moore (who began his career as a Soviet specialist) to reformist liberals like Albert Hirschman (who made his name as a Latin Americanist). Even within the area specialization most closely connected to Cold War concerns—Sovietology itself—there is no discernable ideological pattern linking an area studies orientation to support of U.S. policy. Indeed, after the collapse of the Soviet Union, one could find both conservatives like Frances Fukuyama and leftists like Jerry Hough criticizing Sovietologists for their failure to apply general political science theory to the USSR—and, on the other hand, both conservatives like Martin Malia and leftists like Stephen Cohen blaming Sovietology for its

overreliance on abstract social science models (Fukuyama 1993; Hough 1997; Malia 1993; Cohen 1999). Finally, if the argument that area studies was dominated by inveterate Cold Warriors fails to hold for the Cold War era itself, it fails completely to explain why there was an “area studies” controversy concerning American foreign policy long before World War II—between proponents of Wilsonian universalism, based on Woodrow Wilson’s deeply held convictions about the possibility of expert political science knowledge of the principles ensuring stable democracy and peace, and “unilateralists” skeptical of the relevance of Wilsonian ideals to non-Western contexts (Thorsen 1988; Knock 1992).³ Nor does the presumed pro-American foreign policy bias of area studies hold after the collapse of the Soviet Union—as any reader of Chalmers Johnson’s recent works on U.S. foreign policy can attest (Johnson 2000).

Even if the accusation that area specialists were unduly affected by Cold War priorities seems unwarranted, many analysts would insist that the mission of area studies has nevertheless become anachronistic in the post-Cold War context. Thus one of the most common critiques of area studies research since 1991 has been the argument that it fails to take into account new social, technological, and cultural trends in an era of increasing “globalization” (Prewitt 1996). From this point of view, area studies must be reinvented to account for the rise in influence of non-state actors ranging from democratic advocacy groups to terrorist cells; the erosion of state boundaries as a result of cross-border trade, migration, epidemics, and media; and the invention of new international communication technologies such as the Internet and the cellular phone. Indeed, precisely these sorts of concerns led both the Ford Foundation and the Social Science Research Council—historically both key funders of area studies programs—to cut their support for “traditional” region-based research in the 1990s (Prewitt 1996; Ford Foundation 1999). If this second line of critique is correct, we should expect to find that area specialists prior to 1991 paid little attention in their research to social influences that flow across formal political jurisdictions, to social movements below the level of the nation-state, or to the impact of modern technology on social interaction. Again, however, even a cursory familiarity with the work of leading area specialists throughout the 20th century makes such claims impossible to sustain.

To begin with, area studies as a whole has arguably been *more*, not less, attuned to international influences on national politics than advocates of quantitative political science approaches, who have often of necessity taken the nation-state as the basic unit of analysis in large-*N* multicountry databases. Indeed, the earliest area specialists in Western academia were scholars of ancient religions and their long-term cultural effects on “civilizations”; for them, mere national boundaries were seen as artificial modern divisions that would fail to re-route the deeper allegiances uniting, say, the Muslim, Western Christian, Orthodox, or Confucian “worlds” (Lambert 2001). Area specialists working within the paradigm of modernization theory, too, hardly took the nation-state for granted; rather, they took it upon themselves to explain precisely why this

form of political organization had emerged out of earlier, more parochial forms of political allegiance (E. Weber 1976; Gellner 1983). Others, such as the Western Europeanist Ernst Haas, examined how the nation-state itself might be transcended through the emergence of supranational systems of interest and loyalty in specific regional contexts (Haas 1958, 1964). Nor did area specialists ever tend to neglect the influence of what today would be called “non-state actors”—as is clear from an examination of classic works on the peasant village and peasant rebellion in Mexico, China, and Malaysia (Foster 1965; Johnson 1962; Scott 1985) or on revolutionary movements and ideologies in France and Russia (Sewell 1980; Tucker 1969). As for the impact of technology, this was always one of the central variables in area studies research, which posited increasing urbanization, industrialization, education, and access to media as crucial variables for understanding social change in various regions of the world (Lerner 1958; Apter 1955, 1963; Lowenthal 1970). The typical finding from this body of work was that technological change generally tends to disrupt traditional communities and identities, to facilitate modern forms of individualism, and to enable new possibilities for collective action among previously marginalized groups—precisely the sorts of social challenges and opportunities highlighted in more recent paeans to the “unprecedented” impact of globalization (Friedman 1999, 2005).

This brings us to the third and most widespread stereotype about area studies research: that it has historically been carried out primarily by scholars resistant to general theoretical concepts and innovative methodologies. Taking into account the entire history of the area studies controversy forces us to reevaluate this accusation as well. Specifically, while the original area specialists of the pre-World War II era were primarily humanists interested in topics such as philology, theology, and development of “ancient civilizations”—a tradition now pejoratively labeled “Orientalism” (Said 1978)—one is hard pressed to find representatives of this approach in the comparative politics subfield in any major political science department over the past five decades or so.⁴ In the immediate post-war period, area studies within the American university system was utterly transformed by its close association with Parsonian structural-functionalism and other variants of modernization theory into an integral part of social science theory-building. Given the widespread acceptance of modernization theory’s parsing of the “necessary” stages of social evolution from agricultural “traditional society” based on communalism and personal loyalties to industrial “modern society” based on individualism and impersonal proceduralism, it became possible to conceive of area studies as an arena for scientific investigation of the specific manifestations of “universal” social processes (Parsons 1951; Rostow 1960). Parsons himself articulated such a goal for area studies explicitly as early as 1948, claiming that just as the study of the human body required collaboration among sciences as diverse as “anatomy, physiology, biochemistry, bacteriology, and even psychology and some of the social sciences,” the study of world regions would provide “a concrete focus for the disciplines of the social sciences and related fields of the humanities and natu-

ral sciences” (quoted in Mitchell 2004: 86).

From the 1950s on, it is hard to find respected comparativists arguing that their countries of specialization are simply “unique” and thus not amenable to comparative theorizing. Instead, the goal of most area specialists working in the broad modernization theory tradition was to chart the specific ways in which local “traditional cultures” in their region had interacted with the inescapable forces of urbanization, industrialization, education, and the spread of mass media to produce either successful, stalled, or failed political modernization and economic development. Thus it is not surprising that many social scientists who began as specialists on the effects of modernization in a particular region or country later began to engage in more explicit comparative theorizing—as in the cases of Daniel Lerner, David Apter, Barrington Moore, Lucian Pye, Robert Putnam, and countless other leading comparativists of the period.

It was only when modernization theory began to lose its hegemonic position in the 1970s and 1980s that studies of “political culture” began to be widely seen as atheoretical. Erstwhile students of Parsons himself, such as Clifford Geertz, became increasingly skeptical of their mentor’s claims that human social evolution followed a single universal pattern, leading them to distance themselves from the project of “general social science theory.” Indeed, Geertz’s essay on “thick description”—probably the work most widely cited by critics interested in demonstrating the resistance of area studies to general theory—hardly represented the mainstream area studies viewpoint; rather, it was written in rebellion against the overly deductive forms of “cultural theorizing” that had been dominant in the social sciences for most of Geertz’s early career (Geertz 1973).⁵ At the same time, the search for a new overarching theoretical paradigm began, and this propelled several prominent area specialists to convert to ascendant approaches ranging from neo-Marxism—as in the case of Barrington Moore, whose first books analyzed the impact of modernization on the Soviet dictatorship (Moore 1950, 1954)—to rational choice theory, as in the case of David Laitin, whose first book dealt with the effects of language policy in the single case of post-independence Somalia (Laitin 1977).

By the end of the 1980s, rational choice theory had become the dominant approach in comparative politics research. As older modernization theorists retired and younger scholars trained in rational choice modeling entered the field during the 1990s, then, the idea that one could build a “general theory” by examining global patterns of political culture fell out of favor—and many distinguished scholars who had built their careers pursuing this goal were implicitly recast as traditional “area specialists.” In the 1970s, for example, the prominent modernization theorist Harry Eckstein had firmly declared himself to be a skeptic about “area studies approaches,” arguing that “the ultimate (perhaps only) task of comparative politics is to find general solutions of general problems that cut across both geographic areas and periods of history” (Eckstein 1975: 200). Yet two decades later, two prominent defenders of area studies could cite the very same Eckstein as someone who “developed a powerful theory about how

authority relations in the family influence forms of democracy, based on an intensive study of Norway”—demonstrating the scientific value of the single-country studies (Hall and Tarrow 2001 [1998]: 100). Thus we see that much of the contemporary debate about whether area studies are “theoretical” enough is, on a deeper level, a struggle over the validity of rival theoretical paradigms.

What is an “Area”?

We have seen that critiques of area studies as driven by Cold War concerns, as irrelevant in an era of globalization, or as inimical to generalizable theory—however persistent and pervasive—are fundamentally without merit. There is, however, a fourth common criticism of area studies approaches that is rather more difficult to discard: namely, that the typical definitions of “areas” utilized to demarcate area studies programs at major U.S. universities are arbitrary, and are therefore likely to hinder scientific progress on topics that require conceptualization and research transcending regional borders. Simply put: does organizing comparative politics research in distinct regional clusters make any scientific sense? If not, then even the most dispassionate, dynamic, and theoretically innovative area studies research is built on a faulty foundation.

Ironically, it was Harry Eckstein himself who initially put this theoretical challenge to area studies most clearly. Area studies programs, he noted, might have been formed for any one of four reasons: to fill voids in our factual knowledge about various parts of the world; to provide “rubrics for interdisciplinary collaboration” among scholars with similar linguistic and historical knowledge; to highlight distinctive problems that arise with particular urgency in particular regions; or because “*the societies and politics of different regions constitute distinctive types*” (Eckstein 1975: 202, emphasis in original). Eckstein argues that only the last of these four reasons can justify the maintenance of separate area studies programs for political science research in the long term, from a purely scientific point of view. Once sufficient empirical knowledge about the various polities and societies of the world is gathered, and the causes of urgent political problems in various regions isolated, only an assumption of typological commonality at some level can justify principles of case selection that usually generate comparisons within a particular “area studies” region (say, India with Pakistan) rather than comparisons that ignore such regional designations (say, India with Brazil). Eckstein’s own conclusion was that the most up-to-date research on regime types showed no real justification for distinguishing the study of “Africa,” “South Asia,” “Southeast Asia,” “East Asia,” and so on. At best, only three “clusters” of regimes—European, “Afro-Asian,” and Latin American—might be justified empirically, based on a comprehensive survey of their typical patterns of state authority (Gurr 1974). Ironically, Eckstein noted, “West Europe,” which was one of the last parts of the world to be organized as a separate “area studies” region in American academia, is actually one of the few geographic zones that might actually merit designation as a “area” from a scientific point of view (Eckstein 1975).

Eckstein’s point is a powerful one. Clearly, there is an unavoidable element of arbitrariness in grouping world polities and societies in distinct regional clusters. While perhaps world geography seems at first glance to justify the separate study of, say, “Africa” or “Latin America”—although even in these cases one finds sharp intraregional divisions such as those between Mediterranean North Africa and the sub-Saharan countries, or between Spanish-speaking and Portuguese-speaking South America—one is hard pressed to find purely geographic reasons for distinguishing among the “Middle East,” “South Asia,” “Southeast Asia,” and “East Asia.” Indeed, certain geographical zones that appear to be “regions” in one historical period can suddenly lose their apparent cohesiveness in another. Thus it is not intuitively obvious why, for example, Tajikistan and Slovenia should still be grouped in the same “postcommunist” regional category, now that the Soviet bloc has disintegrated (Rupnik 1999). In the face of such difficulties, shouldn’t we discard the notion of coherent “areas” in world politics altogether—perhaps searching instead for ways to categorize countries that do not depend on the use of “proper nouns” at all (Przeworski and Teune 1970)?

If political scientists already possessed a consensual typology of political regimes as well as a widely accepted general theory of how such regimes develop, rise, and fall, such a line of criticism would, I think, be devastating. Scientific consensus of this sort might render the study of arbitrarily defined “world regions” quite anachronistic. In the absence of such scientific consensus, however, the case for abandoning area studies becomes much less clear cut—and this is a point generally missed by critics of area studies research. Indeed, given the lack of a widely accepted theory of regime evolution in the field, we cannot in principle develop any clear-cut tests of whether particular polities “belong together” in a single regional grouping or not—since only such a theory would allow us to identify precisely those empirical factors that might be decisive for arriving at a scientific judgment in this respect. If so, any provisional “theoretical” categorization of regimes might ultimately turn out to be just as arbitrary as “traditional” area studies divides—if not more so.

Note, for instance, that the tests of regional commonality developed by Eckstein and Gurr depend directly on the theoretical notion that “structures of political authority” (understood in Eckstein’s sense as built around five key dimensions: modes of recruitment, constraints on decision making, levels of participation, intensity of regulation, and complexity) are the central defining feature of all political systems (Gurr 1974; Eckstein and Gurr 1975). If one rejects this theoretical approach—as would most political scientists today—their subsequent grouping of world polities into European, Latin American, and Afro-Asian clusters based on indicators of authority structures no longer has any theoretical or empirical justification. Indeed, choosing another theoretical criterion to generate an empirical categorization of regime types must necessarily generate an entirely different set of “world regions.” Perhaps such alternative theoretical approaches might reveal distinct differences among African, Middle East-

ern, and Asian “clusters” of countries after all—in which case, defenders of traditional “area studies” centers would turn out to have been on the right track, while supporters of “grand theory” might inadvertently have missed crucial regional variables affecting the fate of global political development.

Critics of area studies programs might reply to this point by suggesting that even if many grand theories of regime-types fail, there is at least a chance that *some* general theoretical approach to regime evolution will eventually succeed. Sticking with established regional groupings of countries, in contrast, cannot possibly generate any grand theory. Thus it is still scientifically preferable to reject area studies approaches *a priori*.

This argument, however, assumes that the degree of scientific arbitrariness in the groupings of countries typical of traditional area studies programs approaches 100%: that is, were we to possess a scientific consensus about the general reasons for the emergence, maintenance, and disappearance of political regimes, we would find no theoretical justification for any of the groupings of countries we find in such programs at present. If we hypothesize that existing area studies groupings are only arbitrary to a degree, however—in other words, that there might turn out to be some justification for traditional area studies boundaries, even after the development of a consensual grand theory of regime change—the case for tossing aside area studies research in favor of newly developed theoretical categorizations is substantially weakened. The scientific decision to embrace or to break with area studies should ultimately turn, therefore, on one’s assessment of just how arbitrary the various boundaries of regional studies programs in Western academia actually are, relative to one’s confidence in the ability of new grand theories to categorize polities in more scientifically fruitful ways.

Institutional Legacies, Diffusion, and *Verstehen*

In fact, based on our current state of knowledge, there are many good social scientific reasons to embrace the traditional area studies groupings now institutionalized on campuses throughout North America and Europe. In particular, area studies programs tend to capture quite well the effects of three causal factors that are attracting greater attention among theorists of comparative political and institutional change: the importance of “legacies of the past”; the effects of institutional and social diffusion among neighboring regimes; and the causal effects of powerful ideological, religious, and cultural interpretations of the world.

First, if there is one obvious reason for the persistence over time of traditional area studies boundaries among “African Studies,” “Middle East Studies,” “South Asian Studies,” and so on, it is a shared conviction among area specialists concerning the crucial importance of history. Most fundamentally, the vast majority of contemporary area studies programs are defined in large part by the history of expansion of one or more empires covering much of the territory they examine. Thus it is not much of an exaggeration to describe contemporary “Middle East Studies” as analyzing the territory of the former Ottoman and Persian Empires; “East Asian Stud-

ies” as examining the territory of the Chinese and Japanese Empires; and “Latin American Studies” as covering the territory of the Spanish and Portuguese Empires. The “former Soviet region” is, of course, the most clear-cut case of all in this respect. Other traditionally defined world areas, such as Africa, South Asia, and Southeast Asia, are more notable for the great diversity of empires that occupied their territories in different historical eras. Yet even in these cases, the direct impacts of foreign conquests by various northern peoples are still at the core of regional self-definition. Of course, most area specialists would also insist that their regional programs are justified even more directly by their coverage of distinct linguistic, religious, and cultural contexts. Yet shared languages and belief systems in different parts of the contemporary world, too, can ultimately be traced to the history of imperial expansion and regime change in ancient times.

As contemporary political scientists probe more deeply into the causal effects of structural and institutional legacies inherited from the past, the designation of various regional studies programs to cover geographical zones shaped profoundly by similar experiences of colonial conquest seems to make increasingly good sense. That the legacies of imperial law, bureaucracy, modes of socioeconomic organization, and official cultures might have profound effects on the options available to contemporary state builders is, of course, no surprise for analysts in the historical institutionalist school of comparative politics (Skocpol 1979; Evans, Rueschemeyer and Skocpol 1985; Steinmo, Thelen, and Longstreth 1992; Mahoney and Rueschemeyer 2003). More recently, however, comparativists from the rational choice tradition have also begun to investigate the causal sequences that lead from previous institutional arrangements bequeathed by past regimes to contemporary incentives structuring individual decision making—leading some to conclude that a reconciliation between historical institutionalism and rational choice institutionalism may soon be in the offing (Bates et al. 1998; Thelen 1999). If so, general comparative theorists of both persuasions may soon develop a renewed interest in area studies work on postimperial development and postcolonial state building (Wilkinson, in process).

A second reason why area studies specialists tend to group themselves in distinct research communities that accept traditional regional boundaries is their often implicit understanding that the countries they study tend greatly to influence each other. What happens in Japan, after all, clearly shapes what happens in South Korea—and vice versa—in ways that are clearly not as significant or salient for nations far from the Asian-Pacific region. Similarly, the prospect of accession to the European Union alike had a profound impact on the political and economic institutions of the countries of Eastern and Central Europe; the *acquis communautaire* cannot be said to have had such a dramatic impact on political change in Brazil or Argentina (Ekiert and Hanson 2003; Vachudova 2005). True, area studies programs may sometimes overemphasize the interactions among polities within traditional regional borders at the expense of examination of the influence of neighboring states located in “different” regions; thus

the influence of Russia on Japan, or of China on post-Soviet Central Asia, sometimes receives short shrift in area studies research. Yet even in these cases, area studies specialists may intuitively grasp social limits on cross-border influences in ways that general theorists of “globalization” miss. The social and political impact of South Asia on the Central Asia states since 1991, for example, has arguably been rather marginal compared to that of the longstanding imperial hegemon in the region, Russia—contrary to the assertions of journalistic accounts about the potential for “radical Islam” to spread north from Afghanistan (Rashid 2002).

Again, the typical intuitions of area specialists about the impact and limits of diffusion across regional borders are backed up by new theories and methodologies about the mechanisms of institutional diffusion in international politics. Of course, theorists of economic development have long highlighted the ways in which proximity to the developed capitalist “core” provides structural advantages largely unavailable to countries in the underdeveloped “periphery” (Wallerstein 1975; Evans 1979). Even those who reject dependency theory, however, increasingly accept that geography plays a key role in the global political economy. We now know, more clearly than ever before, just how important geographic “neighbor effects” can be for successful democratization and marketization: in a nutshell, being surrounded by impoverished autocracies makes liberal capitalism extremely difficult, while bordering the European Union in the 1990s nearly guaranteed democracy and wealth (Gallup, Sachs, and Mellinger 1999; Kopstein and Reilly 2000). And where states break down entirely, non-state networks operating in neighboring or nearby states may become the principal conduits of both resources and ideas to would-be revolutionary elites—as we have discovered again in post-invasion Iraq.

This brings us to the third reason why area studies programs tend to maintain their traditional boundaries: namely, despite all of the trends toward cultural globalization over the past two centuries, powerful new ideas about politics must always initially be expressed in some local idiom. Interpretative understanding of such ideas—what Max Weber called the *verstehen* approach—requires a fair degree of contextual knowledge of particular languages and cultures (Weber 1978). Language groupings alone thus provide one clear rationale for traditional area studies programs, since notwithstanding the enormous linguistic diversity of all world regions, there are still major languages in each region that are widely shared among elites—and new political ideas formulated in these languages therefore tend to diffuse quickly along the lines of dominant linguistic communities. Thus fluency in Chinese, Japanese, or Korean is crucial for East Asianists but not for Africanists; fluency in Swahili is a big advantage for a specialist on East Africa but not for a South Asianist. Beyond these obvious linguistic reasons for grouping major world regions, however, is a more subtle cultural one: namely, that shared historical legacies and geographic experiences make particular forms of communication more symbolically meaningful in some regions than others. To understand the political power of ideas like those of Sayyid Qutb in the contempo-

rary Middle East, for example, one needs not only fluency in Arabic, but also enough historical knowledge to see how the common Muslim Arab experience of geopolitical and cultural marginalization since the fall of the Ottoman Empire makes credible a comprehensive theory of global redemption through Islamic jihad against the West (Musallam 2005).

As with the impacts of historical legacies and geographic diffusion, these traditional concerns of “area studies” are largely in tune with cutting edge approaches in the political science discipline. On one hand, historical institutionalists are increasingly interested in charting how the impact of “ideas” may systematically affect the later course of institutional development among countries in a given region (Berman 1996; Hanson 1997; Blyth 2002). On the other hand, rational choice theorists are actively seeking to incorporate the role of “culture”—understood as shared belief systems—into formal models of strategic interaction, and thus accounting for divergent paths of political and economic change among neighboring peoples in otherwise comparable environments (Greif 2006). Of course, even Bates admits that area studies specialists have a strategic advantage over deductive theorists concerning the issue of *verstehen* (Bates 1996). As cutting-edge political science theory illustrates just how much can turn on the cultural meanings attached to particular modes of conduct in different parts of the world, however, that strategic advantage may well begin to count for a great deal more than advocates of deductive theory in the political science discipline initially recognized.

Thus we see that the traditional divisions among major area studies programs, while obviously arbitrary in the sense that their initial organization does not reflect any clear theoretical principle, nevertheless can be justified in ways that are remarkably consistent with emerging theories and methodologies in contemporary comparative politics. Area studies specialists tend to group together polities and societies that have common historical origins in imperial conquest and postcolonial liberation, that geographically occupy the same zones of the world economy and are in a position to influence each other profoundly through institutional diffusion, and that have enough linguistic and cultural commonality to generate distinctive shared responses to new political ideas. Given what we are learning about the importance of past regime legacies, geographical proximity, and cultural interpretations for the course of institutional development in human history, these rough-and-ready principles for the training of future comparativists utilized in most area studies programs may actually serve the comparative politics subfield rather well. If so, discarding the hard-won expertise in regional histories, patterns of economic interaction, languages and cultures built up in area studies programs over the past several decades in favor of newly proposed theoretical categorizations of regimes seems like a very risky bet.

Conclusion

The thrust of this essay has been to provide a defense of the contribution of area studies to comparative politics. As we have seen, most of the common critiques of area studies

are based on incorrect and/or outdated understandings of work in this tradition. And even the strongest theoretical critique of area studies research—that it is built around regional groupings of countries that are in some basic sense arbitrary—can be countered by noting recent theoretical breakthroughs in comparative politics that reemphasize the importance of history, geography, and culture for explaining institutional change.

Returning to the vituperative debate with which this essay began: does this mean that Johnson is right and Bates is wrong? Not entirely. As we have seen, the issues involved in judging the contribution of area studies from a social scientific point of view are extremely complex, and neither side in this longstanding debate has the right to dismiss the contributions of the other out of hand. Indeed, one reason that the debate between area specialists and their critics has continued in much the same form for many decades now is the lack of intellectual respect shown by partisans on both sides, which tends to perpetuate longstanding stereotypes of the “enemy.”

Despite the reasonably large degree of overlap between the criteria for defining “areas” in most area studies programs and the factors now increasingly thought to be vital for explaining patterns of institutional evolution in time and space, there is clearly still some truth to the claim that traditional area studies boundaries are defined in ways that are theoretically and empirically arbitrary. Even if we accept the importance of past legacies of Soviet rule for structuring post-Soviet politics and societies, does the impact of Leninism still justify grouping Tajikistan and Slovenia in the same typological category—two decades after the collapse of communism in East-Central Europe? Or does the divide between the European Union and post-Soviet “Eurasia” now provide a more logical point of departure for comparative political analysis? Similarly, is there really any theoretical reason to study Australia and Burma as common members of “Southeast Asia”? Or should we study Australia and New Zealand as “former British settler states” and thus as decisively different from other postcolonial settings in that region? Concerning such questions, “the facts” will never speak for themselves. Only deductive theory, in the end, can provide clear conceptual criteria for deciding which comparisons will be fruitful for examining particular social scientific hypotheses.

Seen in this light, knowledge of current trends in deductive social science theory can play a crucial intellectual role for specialists in area studies, provoking them constantly to reexamine their understandings of the “regions” on which they claim expertise, and undermining implicitly static characterizations of the nature of regional politics, economics, and cultures. In some cases, longstanding institutional divisions among different area studies centers on American campuses might even be fruitfully rethought in light of critiques developed by more general theories of political and social change. On the other hand, purely deductive analysis, unaccompanied by inductive research into the distinct empirical patterns of human interaction typical of particular parts of the world, will frequently be blind to the ways in which existing historical legacies and geographical diffusion generate the “scope

conditions” that limit the generalizability of most (if not all) political science theories (Hanson and Kopstein 2005). Indeed, a strong training in area studies tends to arm political scientists with precisely the tools they need to see just where “universal” theories of human behavior break down when applied outside the social contexts in which they were first developed. Thus, an open-minded and mutually respectful dialogue between scholars with a preference for deductive models and highly abstract theories, and those who prefer to chart the mechanisms of political and social change that are distinct to particular world areas, might greatly advance our collective understanding of the human condition.

Notes

¹ A longer version of this essay will appear in Todd Landman and Neil Robinson, eds. *Sage Handbook of Comparative Politics*, forthcoming 2009. I would like to thank Margaret Levi, Joel Migdal, and Michael Ward for feedback on a previous version.

² Indeed, in an era when the world’s sole superpower could find only six fluent speakers of Arabic to hire for work in the U.S. Embassy in post-invasion Iraq, one might argue for much greater support for and utilization of area studies expertise in the formulation of U.S. foreign policy (Baker and Hamilton 2006: 92).

³ Note, for example, Theodore Roosevelt’s critique of Wilson’s proposals for the League of Nations as requiring the United States to declare war “every time a Yugoslav wishes to slap a Czechoslovak in the face.” Quoted in Ruggie (1997).

⁴ Perhaps the only prominent exception is Samuel Huntington (1996)—who, ironically, is certainly not known in the field as an “area specialist.”

⁵ One might also point out that Geertz was an anthropologist, rather than a political scientist; hence it hardly makes sense to cite his work as an example of “area studies” methodology in the latter discipline.

References

- Apter, David Ernest. 1955. *The Gold Coast in Transition*. Princeton: Princeton University Press.
- Apter, David Ernest. 1963. *Ghana in Transition*. New York: Atheneum.
- Baker, James A. (III.), Lee H. Hamilton, et al. 2006. *The Iraq Study Group Report*. New York: Vintage Books.
- Bates, Robert H. 1996. “Letter from the President: Area Studies and the Discipline.” *APSA-CP (Newsletter of the APSA Comparative Politics Section)* 7:1 (Winter), 1.
- Bates, Robert H. et al. 1998. *Analytic Narratives*. Princeton, NJ: Princeton University Press.
- Berman, Sheri. 1995. *The Social Democratic Moment: Ideas and Politics in the Making of Interwar Europe*. Cambridge: Harvard University Press.
- Blyth, Mark. 2002. *Great Transformations: Economic Ideas and Institutional Change in the Twentieth Century*. Cambridge: Cambridge University Press.
- Cohen, Stephen. 1985. *Rethinking the Soviet Experience: Politics and History Since 1917*. Oxford: Oxford University Press.
- Cohen, Stephen. 1999. “Russian Studies Without Russia.” *Post-Soviet Affairs* 15:1, 37–55.
- Eckstein, Harry. 1975. “A Critique of Area Studies from a West European Perspective.” In Lucian W. Pye, ed., *Political Science and Area Studies: Rivals or Partners?* (Bloomington: Indiana University Press), 199–217.

- Eckstein, Harry and Ted Robert Gurr. 1975. *Patterns of Authority: A Structural Basis for Comparative Inquiry*. New York: Wiley.
- Ekiert, Grzegorz and Stephen E. Hanson. 2003. "Time, Space, and Institutional Change in Central and Eastern Europe." In Grzegorz Ekiert and Stephen E. Hanson, eds., *Capitalism and Democracy in Central and Eastern Europe: Assessing the Legacy of Communist Rule*. (Cambridge: Cambridge University Press), 15–48.
- Evans, Peter B. 1979. *Dependent Development: The Alliance of Multinational, State, and Local Capital in Brazil*. Princeton: Princeton University Press.
- Evans, Peter B., Dietrich Rueschemeyer, and Theda Skocpol, eds. 1985. *Bringing the State Back In*. Cambridge: Cambridge University Press.
- Farish, Matthew. 2005. "Archiving Areas: The Ethnogeographic Board and the Second World War." *Annals of the Association of American Geographers* 95:3, 663–79.
- Farr, James and Raymond Seideman, eds. 1993. *Discipline and History: Political Science in the United States*. Ann Arbor: University of Michigan Press.
- Ford Foundation. 1999. *Crossing Borders: Revitalizing Area Studies*. New York: Ford Foundation.
- Foster, George M. 1965. "Peasant Society and the Image of Limited Good." *American Anthropologist* 67:2, 293–315.
- Friedman, Thomas L. 1999. *The Lexus and the Olive Tree*. New York: Farrar, Straus, and Giroux.
- Friedman, Thomas L. 2005. *The World is Flat: A Brief History of the Twenty-First Century*. New York: Farrar, Straus, and Giroux.
- Fukuyama, Francis. 1993. "The Modernizing Imperative: The USSR as an Ordinary Country." *National Interest* 31, 10–18.
- Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger. 1999. "Geography and Economic Development." *International Regional Science Review* 22:2, 179–232.
- Geertz, Clifford. 1973. "Thick Description: Toward an Interpretive Theory of Culture." In *The Interpretations of Cultures: Selected Essays*. New York: Basic Books.
- Gellner, Ernest. 1983. *Nations and Nationalism*. Oxford: Blackwell.
- Golden, Miriam. 1998. "Is There an International Division of Labor in Contemporary Political Science: Editor's Introduction." *APSA-CP (Newsletter of the APSA Comparative Politics Section)*, 9:2 (Summer), 6.
- Greif, Avner. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge: Cambridge University Press.
- Gunnell, John G. 1993. "American Political Science, Liberalism, and the Invention of Political Theory." In James Farr and Raymond Seideman, eds., *Discipline and History: Political Science in the United States* (Ann Arbor: University of Michigan Press), 179–97.
- Gurr, Ted Robert. 1974. "Persistence and Change in Political Systems, 1800–1971." *American Political Science Review* 68:4 (December), 1478–81.
- Haas, Ernst B. 1958. *The Uniting of Europe: Political, Social, and Economic Forces, 1950–1957*. Stanford: Stanford University Press.
- Haas, Ernst B. 1964. *Beyond the Nation-State: Functionalism and International Organization*. Stanford: Stanford University Press.
- Hall, Peter A. and Sidney Tarrow. 2001 (1998). "Globalization and Area Studies: When is Too Broad Too Narrow?" In Patrick O'Meara, Howard D. Mehlinger, and Roxana Ma Newman, eds., *Changing Perspectives on International Education* (Bloomington: Indiana University Press, 2001), 96–100.
- Hanson, Stephen E. 1997. *Time and Revolution: Marxism and the Design of Soviet Institutions*. Chapel Hill: University of North Carolina Press.
- Hanson, Stephen E. and Jeffrey S. Kopstein. 2005. "Regime Type and Diffusion in Comparative Politics Methodology." *Canadian Journal of Political Science* 38 (March), 69–99.
- Hough, Jerry F. 1997. *Democratization and Revolution in the USSR*. Washington, DC: Brookings.
- Huntington, Samuel P. 1996. *The Clash of Civilizations and the Remaking of World Order*. New York: Simon and Schuster.
- Johnson, Chalmers. 1982. *Peasant Nationalism and Communist Power: The Emergence of Revolutionary China, 1937–1945*. Stanford: Stanford University Press.
- Johnson, Chalmers. 2000. *Blowback: The Costs and Consequences of American Empire*. New York: Metropolitan Books.
- Johnson, Chalmers. 2001. "Preconception vs. Observation, or the Contributions of Rational Choice Theory and Area Studies to Contemporary Political Science." In Patrick O'Meara, Howard D. Mehlinger, and Roxana Ma Newman, eds., *Changing Perspectives on International Education* (Bloomington: Indiana University Press), 58–66.
- Knock, Thomas J. 1992. *To End All Wars: Woodrow Wilson and the Quest for a New World Order*. New York: Oxford University Press.
- Kopstein, Jeffrey and David Reilly. 2000. "Geographic Diffusion and the Transformation of the Postcommunist World." *World Politics* 53 (October), 1–37.
- Laitin, David D. 1977. *Politics, Language, and Thought: The Somali Experience*. Chicago: University of Chicago Press.
- Lambert, Richard D. et al. 1984. *Beyond Growth: The Next Stage in Language and Area Studies*. Washington, DC: Association of American Universities.
- Lambert, Richard D. 2001. "Domains and Issues in International Studies." In Patrick O'Meara, Howard D. Mehlinger, and Roxana Ma Newman, eds., *Changing Perspectives on International Education* (Bloomington: Indiana University Press, 2001), 30–48.
- Lerner, Daniel. 1958. *The Passing of Traditional Society: Modernizing the Middle East*. Glencoe, IL: Free Press.
- Lowenthal, Richard. 1970. "Development v. Utopia in Communist Policy." In Chalmers Johnson, ed., *Change in Communist Systems* (Stanford: Stanford University Press), 33–116.
- Mahoney, James and Dietrich Rueschemeyer, eds. 2003. *Comparative Historical Analysis in the Social Sciences*. Cambridge: Cambridge University Press.
- Malia, Martin. 1993. "A Fatal Logic." *National Interest* 31: 80–90.
- Mitchell, Timothy. 2004. "The Middle East in the Past and Future of Social Science." In David Szanton, ed., *The Politics of Knowledge: Area Studies and the Disciplines*. (Berkeley: University of California Press), 74–118.
- Moore, Barrington. 1950. *Soviet Politics: The Dilemma of Power*. Cambridge: Harvard University Press.
- Moore, Barrington. 1954. *Terror and Progress USSR: Some Sources of Change and Stability in the Soviet Dictatorship*. Cambridge: Harvard University Press.
- Musallam, Adnan. 2005. *From Secularism to Jihad: Sayyid Qutb and the Foundations of Radical Islamism*. Westport, CT: Praeger.
- Parsons, Talcott. 1951. *The Social System*. Glencoe, IL: Free Press.
- Prewitt, Kenneth. 1996. "Presidential Items." *Items* 50:1 (March): 15–18.
- Przeworski, Adam and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience.
- Rashid, Ahmed. 2002. *Jihad: The Rise of Militant Islam in Central Asia*. New Haven: Yale University Press.
- Rostow, W.W. 1960. *The Stages of Economic Growth: A Non-Communist Manifesto*. Cambridge: Cambridge University Press.
- Ruggie, John Gerard. 1997. "The Past as Prologue?: Interests, Identity, and American Foreign Policy." *International Security* 21:4, 89–125.

- Rupnik, Jacques. 1999. "The Postcommunist Divide." *Journal of Democracy* 10:1, 57–62.
- Said, Edward W. 1978. *Orientalism*. New York: Vintage Books.
- Scott, James C. 1985. *Weapons of the Weak: Everyday Forms of Peasant Resistance*. New Haven: Yale University Press.
- Sewell, William. 1980. *Work and Revolution in France: The Language of Labor from the Old Regime through 1848*. Cambridge: Cambridge University Press.
- Skocpol, Theda. 1979. *States and Social Revolutions*. Cambridge: Harvard University Press.
- Steinmo, Sven, Kathleen Thelen, and Frank Longstreth, eds. 1992. *Structuring Politics: Historical Institutionalism in Comparative Analysis*. Cambridge: Cambridge University Press.
- Thelen, Kathleen. 1999. "Historical Institutionalism in Comparative Politics." *Annual Review of Political Science* 2, 369–404.
- Thorsen, Neils. 1998. *The Political Thought of Woodrow Wilson, 1875–1910*. Princeton: Princeton University Press.
- Tucker, Robert C. 1969. *The Marxian Revolutionary Idea*. New York: Norton.
- Vachudova, Milada Anna. 2005. *Europe Undivided: Democracy, Leverage, and Integration after Communism*. Cambridge: Cambridge University Press.
- Wallerstein, Immanuel. 1974. "The Rise and Future Demise of the World Capitalist System: Concepts for Comparative Analysis." *Comparative Studies in Society and History* 16:4 (September), 387–415.
- Ward, Robert E. 1975. "Culture and the Comparative Study of Politics." In Lucian W. Pye, ed., *Political Science and Area Studies: Rivals or Partners?* (Bloomington: Indiana University Press), 23–47.
- Weber, Eugene Joseph. 1976. *Peasants into Frenchmen: The Modernization of Rural France, 1870–1914*. Stanford: Stanford University Press.
- Weber, Max. 1978. *Economy and Society: An Outline of Interpretive Sociology*. Berkeley: University of California Press.
- Wilkinson, Steven I. In process. "Colonization, Institutions and Conflict."

Book Notes

Book descriptions are excerpted from publisher's websites. If you would like to recommend a book to be included in this section, email Joshua C. Yesnowitz, the assistant editor of QM, at jcyesnow@bu.edu.

- Firebaugh, Glenn. 2007. *Seven Rules for Social Research*. Princeton: Princeton University Press.

Seven Rules for Social Research teaches social scientists how to get the most out of their technical skills and tools, providing a resource that fully describes the strategies and concepts no researcher or student of human behavior can do without. Glenn Firebaugh provides indispensable practical guidance for anyone doing research in the social and health sciences today, whether they are undergraduate or graduate students embarking on their first major research projects or seasoned professionals seeking to incorporate new methods into their research. The rules are the basis for discussions of a broad range of issues, from choosing a research question to inferring causal relationships, and are illustrated with applications and case studies from sociology, economics, political science, and related fields. Though geared toward quantitative methods, the rules also work for qualitative research. The Seven Rules are: (1) there should be the possibility

of surprise in social research; (2) look for differences that make a difference, and report them; (3) build reality checks into your research; (4) replicate where possible; (5) compare like with like; (6) use panel data to study individual change and repeated cross-section data to study social change; and (7) let method be the servant, not the master.

- Kenworthy, Lane and Alexander Hicks, eds. 2008. *Method and Substance in Macrocomparative Analysis*. London: Palgrave.

Macrocomparative researchers use a variety of methodological approaches. This book features analyses of a single substantive topic, comparative employment performance in affluent countries, using three of the most common macrocomparative techniques: pooled cross-section time-series regression, qualitative comparative analysis (QCA), and small-*N* analysis. The chapters assess the strengths and weaknesses of these three approaches. They also examine important questions of causal inference, such as selection bias. Contributors include prominent methodologists and comparative researchers from political science and sociology.

- Parsons, Craig. 2007. *How to Map Arguments in Political Science*. Oxford: Oxford University Press.

In order to venture into explanations of political action we need a map of our basic options: what types of examinations are out there? Even advanced students and scholars can find the landscape difficult to navigate. They confront a bewildering maze of partial typologies, contrasting uses of terms, and debates over what counts as an explanation. *How to Map Arguments in Political Science* provides a basic map and toolkit for analysis in political science that references political examples and a wide range of material from the political science literature. While common terms—structural, institutional, ideational, and psychological logics—form the sectors of the map, this book is unique in its arguments regarding how to best define these terms. Adopting a systematic and exhaustive framework, it defines the main analytical approaches in ways that facilitate both competition and combination. Finally, it leads to revisions of prevailing views on philosophy of science and research design to encourage more open and rigorous debates.

Article Notes

- Ashworth, Scott, Joshua D. Clinton, Adam Meirowitz, and Kristopher W. Ramsay. 2008. "Design, Inference, and the Strategic Logic of Suicide Terrorism." *American Political Science Review* 102:2 (April) 269–73.

In "The Strategic Logic of Suicide Terrorism," Robert Pape (2003) presents an analysis of his suicide terrorism data. He uses the data to draw inferences about how territorial occupation and religious extremism affect the decision of terrorist groups to use suicide tactics. We show that the data are incapable of supporting Pape's conclusions because he "samples on the dependent variable." (The data only contains cases in which suicide terror is used.) We construct bounds (Manski 1995) on the quantities relevant to Pape's hypotheses and show exactly how little can be learned about the relevant statistical associations from the data produced by Pape's research design.

Blatter, Joachim and Till Blume. 2008. "In Search of Co-variation, Causal Mechanisms or Congruence? Towards a Plural Understanding of Case Studies." *Swiss Political Science Review* 14:2 (Summer) 315–55.

Methodological reflections about case study research have increased within recent years. According to our account, there are three distinct approaches to case studies: co-variational, causal process tracing, and congruence analysis. The main goals of this article are to lay out the distinct ways in which causal inferences are drawn for the cases under study and to scrutinize the different understandings and directions of generalization within these three approaches. By doing so we highlight two aspects: First, causal process tracing and congruence analysis should be seen as two distinct alternatives to the dominant co-variational template. Second, the main characteristics of case studies, their thickness, provides only an unavoidable dilemma if we aim to generalize the findings towards a wider population of similar cases as in the co-variational template. If we would like to get deeper insights—as the causal process tracing approach does—or if we would like to use the empirical evidence for a broader theoretical discourse—as the congruence analysis does—case studies do not face a tradeoff.

Chandra, Kanchan and Steven Wilkinson. 2008. "Measuring the effect of 'ethnicity'." *Comparative Political Studies* 41:4–5, 515–63.

Most tests of hypotheses about the effects of "ethnicity" on outcomes use data or measures that confuse or conflate what are termed *ethnic structure* and *ethnic practice*. This article presents a conceptualization of ethnicity that makes the distinction between these concepts clear; it demonstrates how confusion between structure and practice hampers the ability to test theories; and it presents two new measures of ethnic practice—ECI (the ethnic concentration index) and EVOTE (the percentage of the vote obtained by ethnic parties)—that illustrate the pay-offs of making this distinction and collecting data accordingly, using examples from the civil war literature.

Clemens, Elisabeth S. 2007. "Toward a Historicized Sociology: Theorizing Events, Processes, and Emergence." *Annual Review of Sociology* 33 (August) 527–49.

Since the 1970s, historical sociology in the United States has been constituted by a configuration of substantive questions, a theoretical vocabulary anchored in concepts of economic interest and rationalization, and a methodological commitment to comparison. More recently, this configuration has been destabilized along each dimension: the increasing autonomy of comparative-historical methods from specific historical puzzles, the shift from the analysis of covariation to theories of historical process, and new substantive questions through which new kinds of arguments have been elaborated. Although the dominant responses have centered on methodological elaboration and epistemological debate, greater attention to historical process also informs new strategies for defining cases and framing puzzles, thereby highlighting different categories of empirical questions: social caging, group formation, and multiple orders. The most striking shift is from the imagery of systems-and-crises, which highlighted revolution and state-building, to multidimensional understandings of emergence and destabilization.

Hyde, Susan D. 2007. "The Observer Effect in International Politics." *World Politics* 60:1 (October), 37–63.

By pressuring governments to hold democratic elections and by becoming directly involved in the electoral process through technical assistance and funding or as election monitors, international actors now play a visible role in domestic elections and other democratic processes throughout the developing world. Although scholars have documented several macrolevel relationships between international-level variables and movement toward democracy, there has been little attention paid to the microlevel effects of international involvement in the democratization process. This article examines the effects of international election observation as a prominent form of international involvement in domestic elections and exploits a natural experiment in order to test whether international observers reduce election fraud. Using data from the 2003 presidential elections in Armenia, the article demonstrates that although observers may not eliminate election fraud, they can reduce election-day fraud at the polling stations they visit. The unusual advantage of experiment-like conditions for this study offers unique causal evidence that international actors can have direct, measurable effects on the level of election-day fraud and, by extension, on the democratization process.

Klein, James P., Gary Goertz, and Paul F. Diehl. 2008. "The Peace Scale: Conceptualizing and Operationalizing Non-Rivalry and Peace." *Conflict Management and Peace Science* 25:1 (March), 67–80.

Most efforts in international relations scholarship focus on understanding war and conflict. To the extent that peace is considered, it is often treated as an afterthought or a control condition. In this paper, we construct a five-level scale of dyadic peace. We then operationalize the scale using indicators suggested by different scholarly literatures as well as measures derived from case studies. Further, we advocate the use of "anchor cases" and pair-wise comparisons to code dyadic peace in non-rivalry periods. We conclude with a pilot study of the scale, as applied to pairs of states over the period from 1816 to 2006.

Shalev, Michael. 2007. "Limits and Alternatives to Multiple Regression in Comparative Research." *Comparative Social Research* 24, 261–308.

This paper criticizes the use of multiple regression (MR) in the fields of comparative social policy and political economy and proposes alternative methods of numerical analysis. The limitations of MR in its characteristic guise as a means of hypothesis-testing are well known. The emphasis here is on the specific difficulties of applying MR to the problem of explaining diverse outcomes across a limited range of country cases. Two principal conclusions will emerge. First, even though technical means are available to deal with many of the limitations of MR, these solutions are either unconvincing or else require such advanced technical skills that they offer questionable returns on scholarly investment. Second, dissatisfaction with MR does not necessarily mandate radical alternatives or abandonment of numerical methods altogether. "Low-tech" forms of analysis (tabular and graphical methods) and multivariate statistical techniques other than MR (such as factor analysis) constitute viable and useful alternatives.

Announcements

Giovanni Sartori Award for the Best Book Published in 2007 Developing or Applying Qualitative Methods

Recipient: Michael Tomz, Stanford University, *Reputation and International Cooperation: Sovereign Debt Across Three Centuries*. Princeton, NJ: Princeton University Press, 2007.

Award Committee: Kristen Renwick Monroe (chair), Amel Ahmed, Fred Chernoff.

Citation: The Committee had many excellent books from which to choose and found itself incredibly impressed, if a little overwhelmed, with the array of talent and methodological sophistication in the discipline. In making our decision, we tried to focus on the best-quality book that fell within the parameters of the Sartori Award itself, which is designed to honor the best book that demonstrates originality and innovation in the use or development of qualitative methods. The Committee looked for works that choose a problem because of its importance and not because it was amenable to straightforward sorts of methods, because it demonstrated an interesting use of methods to analyze the data relevant for the topic and because the breadth of topic gave it political and intellectual significance. Since the Sartori award was established to honor works utilizing and developing qualitative methods, works that did a good job methodologically but which relied primarily on quantitative methods were ranked somewhat lower; this decision does not suggest such works or methodological approaches are deficient but instead acknowledges the Award's charge to highlight innovating works using qualitative methodology. The award also encompasses works that probe issues in the philosophy of science and social science and touches on how debates in those areas affect substantive research.

The winner of the 2008 Giovanni Sartori Award is Michael Tomz of Stanford University for *Reputation and International Cooperation: Sovereign Debt Across Three Centuries*. In this book, Tomz seeks to explain how credibility on international debt can be sustained in an anarchical world—in other words, why investors lend and governments repay. To this end he culls data on international capital markets from the past 300 year to test his reputational theory. The 300 year span is ambitious and presents some daunting methodological challenges. The study is impressive in the breadth and creativity of the methodological approach, combining archival material, case studies, and where possible quantitative data to develop a “reputational theory of cooperation.” Tomz' work is exemplary in how it applies a wide and eclectic array of methods in the service of problem-driven research. He applies these tools systematically and rigorously to produce a fascinating study and important contribution.

Honorable mentions go to Gary Goertz (University of Arizona) and Jack Levy (Rutgers University) for *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*. (Routledge Press, 2007). This is an edited volume with an impressive collection of essays that survey the ways in which necessary conditions counterfactuals are used in the logics of causal explanation, particularly in explanations of international conflict. The work represents an important contribution to the study of methods in general and usefully develops some central concepts in qualitative research. The substantive focus also helps to elucidate the more abstract theoretical discussions.

Honorable mentions go to Craig Parsons (University of Oregon) for *How to Map Arguments in Political Science* (Oxford University Press, 2007). In this book Parsons takes on the daunting task of organizing the epistemological and ontological commitments of po-

litical science scholarship. This is very useful contribution and also a very clear and well-designed book that does an excellent job of identifying the distinctive features of different types of argument and explanations.

Alexander George Award for the Best Article or Book Chapter Published in 2007 Developing or Applying Qualitative Methods

Recipients: Giovanni Capoccia, Oxford University, and R. Daniel Kelemen, Rutgers University, “The Study of Critical Junctures: Theory, Narrative, and Counterfactuals in Historical Institutionalism.” *World Politics* 59:3.

Award Committee: Bear Braumoeller (chair), Cynthia Kaplan, George Thomas.

Citation: In many ways, Capoccia and Kelemen's article exemplifies the tradition that Alexander George pioneered. It addresses an issue that is very often mentioned in passing but is much less often taken up with great care, and it deals with it thoughtfully, systematically, and well. It very substantially improves on the state of the art in thinking about its subject. Finally, and most importantly, it is useful: it holds forth great promise of making other people's work considerably better. We congratulate the authors on a job well done.

SAGE Prize for the Best Paper Presented at the 2007 Annual Meeting of the American Political Science Association Developing or Applying Qualitative Methods

Recipients: Dan Slater and Erica Simmons, “Informative Regress: Critical Antecedents in Comparative-Historical Analysis”

Award Committee: Frederic Schaffer (chair), Sarah Lischer, Michael Coppedge.

Citation: Slater and Simmons make a strong and lively argument that comparative-historical researchers should pay more explicit and systematic attention to what they call “critical antecedents”—conditions predating a critical juncture that “combine with causal factors during a critical juncture to produce divergent outcomes.” Critical junctures, the authors teach us, are parts of longer causal chains. By taking a longer view, we not only discover that factors predating the critical juncture can still be an important part of an explanation, we are also more apt to notice areas of agreement between explanations that seemingly compete. Slater and Simmons thus make an innovative and constructive critique of the prevailing wisdom on comparative-historical analysis. Their paper is also a model of clear exposition. By examining carefully the arguments developed in a handful of well-chosen books, the authors show in a concrete and engaging way the payoff of taking their approach seriously.

Comparative Political Studies Announces new “Methodology Forum”

Comparative Political Studies is pleased to announce a new feature of our journal, namely a “Methodology Forum” in which we hope to print methodological articles with a comparative scope and occasional exchanges on important issues of comparative methodology. While the Forum will run on an occasional basis, we do hope to encourage submissions of both quantitative and qualitative articles with a methodological message for comparative politics scholars.

Announcing the Eighth Institute for Qualitative and Multi-Method Research, Syracuse University, May 25–June 10, 2009

The Consortium on Qualitative Research Methods (CQRM) is pleased to announce the eighth Institute for Qualitative and Multi-Method Research.

The institute seeks to enable students to create and critique methodologically sophisticated qualitative research designs, including case studies, tests of necessity or sufficiency, and narrative or interpretive work. It will explore the techniques, uses, strengths, and limitations of these methods, while emphasizing their relationships with alternative approaches. Topics include research design, concept formation, methods of structured and focused comparisons of cases, typological theory, case selection, process tracing, comparative historical analysis, congruence testing, path dependency, interpretivism, counterfactual analysis, interview and field research (including archival) techniques, necessary and sufficient conditions, fuzzy set methods, and philosophy of science issues relevant to qualitative research.

Attendees will receive constructive feedback on their own qualitative research designs, and the course will also include master class discussions led by the authors of well known works which employ qualitative methods. Examples will be drawn from exemplary research in international relations, comparative politics, and American politics. The schedule and reading list from the seventh annual institute, available on the CQRM website at <http://www.maxwell.syr.edu/moynihan/programs/cqrm/institute.html>, indicates the range of the issues to be covered. Please note, however, that this syllabus will be revised for the eighth institute, and should be viewed with this in mind.

CQRM member institutions will use their own meritocratic criteria to select students or junior faculty to attend the institute, and should notify CQRM of their choices by February 13, 2009. Students, fellows and junior faculty who are not sure if they will be selected, or who attend non-member organizations, should apply directly to CQRM using the form available at this link: <http://survey.maxwell.syr.edu/Survey.aspx?s=5bfb265123954e9eb2a7cbbf46d13707>. Open pool applications must be received by November 15, 2008. Applicants will be notified of the outcome by December 9, 2008.

CQRM will cover the costs of tuition, lodging, and meals for successful applicants. Attendees will be responsible for their own transportation costs to and from Syracuse University. Participants for the institute will arrive on Monday, May 25, and depart late Tuesday June 9, or any time on Wednesday June 10, 2009. The seminar will meet daily, beginning on Tuesday, May 26, with some break days to be announced. The final meeting is scheduled for Tuesday, June 9.

The 2009 Institute for Qualitative and Multi-Method Research will be the first to be held at the Maxwell School, Syracuse University, and consequently to be scheduled for a summer time slot. We anticipate that the move to Maxwell will bring several advantages. First, the Moynihan Institute and the Maxwell School are providing substantial long-term support for further broadening our activities. Second, the institute was previously bracketed by New Years Day and the beginning of the Arizona State University Spring semester, limiting us to two weeks. By moving to the summer, we will be able to have a longer and more flexible timetable. Third, in recent years we have seen a very substantial increase in participation by attendees from European universities, and this move significantly reduces the travel distance for those students. Finally, the move to a venue with the standing and reputation of the Maxwell School is further recognition of the institute's growing national and international prominence.

On behalf of the Consortium, and of the 630 graduate students and faculty who attended the institute when it was located in Tempe, we want to take this opportunity to thank Arizona State University, and especially the Department of Political Science, for its outstanding support.

Announcing a New APSA Conference Group on Interpretive Methodologies & Methods

A new APSA Conference Group on Interpretive Methodologies & Methods has been formed to provide a forum for the discussion of methodologies and methods related to interpretive research.

Widely used across the full range of the social sciences and building on over 100 years of theoretical writings, interpretive research methods are informed, explicitly or implicitly, by presuppositions deriving from phenomenology, hermeneutics, and some critical theory, of European (Continental) background, and pragmatism, symbolic interaction, and ethno-methodology, developed in the US. Their concerns often overlap with such other approaches as feminist theories, critical race theory, and critical legal studies. Although diverse in their modes of accessing/generating and analyzing data, research processes in the constructivist-interpretive traditions are united by an empirical and normative prioritizing of the lived experience of people in research settings (what the anthropologist Clifford Geertz referred to as "experience-near" research, also known as "emic" research), a focus on the meaning(s) of acts, events, interactions, language, and physical artifacts to multiple stakeholders, and the potential for plurality in sense-making of those artifacts, leading at times to conflict.

Interpretive scholars also diverge with respect to the ends they pursue: they may be practical or contemplative, they may seek to educate, they may be concerned with the pursuit of the particular or with a grand sweeping synthesis, and so on. For some, the legitimation of local knowledge constitutes a radically democratic position; for some, the work requires a commitment to individual human agency; and for yet others it serves to challenge the modern constitution of the subject.

The Conference Group will provide a forum for these and other discussions. An important motivation behind its formation has been the need to make space for interpretive methodology as a legitimate approach within US political science. All political scientists share a desire to see, understand, and explain the social world. The systematic inclusion of interpretive approaches (in graduate curricula, disciplinary journals, research funding programs, and so on) carries enormous potential to enliven and enrich our discipline's current ways of seeing, understanding, and explaining. The work showcased in this Conference Group will highlight how much political science stands to gain from the wider inclusion of interpretive approaches.

Some of the interesting interpretive work being done today originates outside the field of political science, but there are also longstanding interpretive traditions within it. We look forward to a constructive dialogue among political scientists interested in interpretive work, whether academics or practitioners (e.g., policy analysts), including those doing research that is interpretive only in part (such as some forms of ethnography). We envision discussions ranging from the practical to the philosophical, including career-related issues (e.g., publishing and/or getting funding for interpretive research, teaching interpretive methods), panels on specific methods and methodological issues, strategies for effectively developing and conducting interpretive work, presentations of interpretive research, and issues in the philosophy of (social) science (such as the question of realism, mixing ontological and epistemological positions, etc.).

While the Qualitative and Multi-Method Research section has provided a forum for interpretive papers and panels since its incep-

tion, the growing community of interpretive researchers recognizes the need for a self-standing group, with “interpretive” in its name, that will be able to give voice to these ideas in panels of its own design, concerning issues that are unique to its concerns. Its distinctive name identifies a specific site in the annual meeting where interpretive scholars can meet those who share these interests, in ways that cannot be sustained through Working Groups and Short Courses. In addition to panels featuring interpretive work, we envision co-sponsored panels with QMMR on topics that are of shared interest—e.g., qualitative and interpretive approaches to research design, concept development, etc.—in ways that help all of us to clarify and understand the similarities and differences; and we hope that the section will be open to such collaboration. We see this Conference Group as an extension of the activities of QMMR, and we will continue to encourage Group members to join the section and attend its panels and other events.

The Call for Papers for APSA 2009 is available at http://apsanet.org/content_58183.cfm.

Palgrave ECPR Research Methods Series

In association with the ECPR Standing Group for Political Methodology, Palgrave Macmillan is delighted to announce a new series dedicated to producing cutting-edge titles in research methods. Palgrave Macmillan has been a publisher of outstanding scholarly work on international affairs since the inception of an academic discipline of IR in the 1930s, with a list that includes many of its seminal thinkers. Palgrave Studies in International Relations will build on that tradition.

Edited by Bernhard Kittel and Benoît Rihoux, the Palgrave ECPR Research Methods Series will provide students and scholars with state-of-the-art scholarship on methodology, methods, and techniques. The series will comprise innovative and intellectually rigorous monographs and edited collections which bridge schools of thought and cross the boundaries of conventional approaches. It will cover both empirical-analytical and interpretive approaches, micro and macro studies, and quantitative and qualitative methods.

The aim of the series is to present research methods in an accessible form, suitable both for researchers and graduate course-takers and tutors working in this broad and rapidly changing field. The presentation of methodology, methods and techniques is always embedded in the context of specific research problems. Explicating the relationship between approach, theory, method, and substantive topic is at the core of the series. Future titles are expected, among others, on the analysis of political attitudes and behavior, on the treatment of time and sequence, on case studies, on the collection, measurement, and analysis of macro data, on policy studies, on experimental research, and on mixed methods in political science.

Titles: *Method and Substance in Macrocomparative Analysis*. Lane Kenworthy and Alexander Hicks, eds. (2008); *Qualitative Methods in International Relations*, Audie Klotz and Deepa Prakash, eds. (2008); *Experts and Elite Interviews: Methodology and Practice*. Alexander Bogner, Beate Littig, and Wolfgang Menz, eds. (forthcoming 2009). To order, access <http://www.palgrave.com/products/Title.aspx?PID=296391>.

For more information, please contact:

Bernhard Kittel, bernhard.kittel@uni-oldenburg.de

Benoît Rihoux, benoit.rihoux@uclouvain.be

Alexandra Webster, A.Webster@Palgrave.com

Chapters Relevant to the Study of Qualitative Methods in Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, eds. *The Oxford Handbook of Political Methodology* (New York: Oxford University Press, 2008)

- Russell Hardin, “Normative Methodology” (Ch. 2)
Mark Bevir, “Meta-Methodology: Clearing the Underbrush” (Ch. 3)
Gary Goertz, “Concepts, Theories, and Numbers: A Checklist for Constructing, Evaluating, and Using Concepts or Quantitative Measures” (Ch. 5)
David Collier, Jody LaPorte, and Jason Seawright, “Typologies: Forming Concepts and Creating Categorical Variables” (Ch. 7)
Charles C. Ragin, “Measurement versus Calibration: A Set-Theoretic Approach” (Ch. 8)
Henry E. Brady, “Causation and Explanation in Social Sciences” (Ch. 10)
David A. Freedman, “On Types of Scientific Enquiry: The Role of Qualitative Reasoning” (Ch. 12)
Peter Hedström, “Studying Mechanisms to Strengthen Causal Inferences in Quantitative Research” (Ch. 13)
Alan S. Gerber and Donald P. Green, “Field Experiments and Natural Experiments” (Ch. 15)
Jack S. Levy, “Counterfactuals and Case Studies” (Ch. 27)
John Gerring, “Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques” (Ch. 28)
Brian C. Rathbun, “Interviewing and Qualitative Field Methods: Pragmatism and Practicalities” (Ch. 29)
Andrew Bennett, “Process Tracing: A Bayesian Perspective” (Ch. 30)
Benoît Rihoux, “Case-Oriented Configurational Research: Qualitative Comparative Analysis (QCA), Fuzzy Sets, and Related Techniques” (Ch. 31)
James Mahoney and P. Larkin Terrie, “Comparative-Historical Analysis in Contemporary Political Science” (Ch. 32)
James D. Fearon and David D. Laitin, “Integrating Qualitative and Quantitative Methods” (Ch. 33)
David Collier and Colin Elman, “Qualitative and Multi-Method Research: Organizations, Publication, and Reflections on Integration”
-

Qualitative and Multi-Method Research

Department of Political Science
315 Social Sciences Building
P.O. Box 210027
University of Arizona
Tucson, AZ 85721-0027
USA

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

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: jyesnow@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.