

BOOK REVIEWS

Symposium on Qualitative Research Methods in Political Science

Colin Elman, Arizona State University

doi:10.1017/S0022381607080206

For this issue of the *Journal of Politics*, several recent volumes on qualitative research methods and related topics are reviewed. This short introduction provides a brief background commentary on recent developments in this field of methodology.

Qualitative research methods are commonly used in political science, especially in the International Relations and Comparative Politics subfields (Bennett and Elman 2007a, 2007b; Mahoney 2007). To some extent, this interest has not been reflected in contemporary studies of American politics, which emphasize statistical analysis and formal modeling (Pierson 2007). Nevertheless, taking political science as whole, one would have to go back to at least the mid-1970s to find such a concentrated burst of developments.

Contemporary scholars working on and with qualitative research methods comprise a large and varied community. Pluralism within the field has been exhibited in the increasingly sophisticated use of multiple methods in composite research designs, through nesting, iteration and other strategies. In addition to this combination of qualitative and other research methods, the qualitative research community itself has outgrown a one-size-fits-all approach and can be seen to encompass a rich diversity of debates.

Among the more significant of these discussions are those concerning the epistemological underpinnings of qualitative analysis. On the one hand, those scholars who follow a single-logic-of-inference model—most notably the adherents of *Designing Social Inquiry* (King, Keohane, and Verba 1994)—would argue for a unity of approach across social science techniques. According to this view, because quantitative and qualitative methods rely on the same logic of inference, they have the same constraints and problems.

By contrast, other proponents of qualitative research methods—including adherents of the ap-

proaches suggested by Brady and Collier (2004) and George and Bennett (2005)—argue that qualitative methods are not quantitative methods writ small. In particular, within-case methods are often deployed to establish the presence or absence of causal mechanisms and the conditions under which they operate (Bennett and Elman 2007c; Mahoney and Goertz 2006). As a result, to the extent they use alternative logics of inference, qualitative research methods are not susceptible to the same potential weaknesses as quantitative approaches, such as often mentioned issues arising from degrees of freedom and case selection (Bennett and Elman 2006a). To be sure, qualitative methods have their own serious inferential problems to avoid, but the multiple-logics-of-inference view is that they also bring unique strengths to the table. For example, causally complex situations may sometimes be best addressed using within-case analysis (Bennett and Elman 2006b).

A second set of concerns arises from qualitative research methods' engagement with constructivist ontologies. Although there is near universal agreement among constructivists that this engagement requires an interpretivist epistemology, there is considerable variation in how scholars approach interpretive tasks. For example, to borrow Ruggie's (1998) tripartite classification, one suspects that there would be a considerable overlap between naturalistic and pragmatic constructivists and methodologists who hold to the multiple-logics-of-inference view. By contrast, postmodernist constructivists are likely to be highly skeptical of allying interpretivism with even this more eclectic branch of qualitative methods.

Despite encompassing a diversity of approaches, the overwhelming majority of the qualitative research community occupies a large and incredibly productive middle ground. There may be a few scholars at the different edges whose epistemological presuppositions require taking an exclusivist approach. Most scholars, however, are comfortable with the pluralist view that different methods have different strengths and weaknesses, and one should choose and use them accordingly. To be sure, there are vibrant and sometimes strong disagreements on any number of topics, but these have

taken place in the context of an increasingly self-aware and well-defined research community.

The growth of interest in qualitative research methods has caused, and in turn been partly caused by, the development of related institutions. The American Political Science Association's (APSA) organized section on Qualitative and Multi-Method Research has approximately 1000 members and is now the third largest of APSA's 37 sections. The section began in 2003 with an annual meeting allocation of 6 panels, and by 2007 it had 26 slots (30 with cosponsorships). At APSA 2007, roughly 200 scholars signed up for short courses and a working group co-organized by the section. A related effort, the annual Arizona State University Institute for Qualitative and Multi-Method Research has been held each January since 2002 and is now supported by more than 60 member departments and funding from the National Science Foundation. Beginning with 45 students in 2002, and growing to approximately 130 in 2008, the institute has trained more than 600 students in advanced qualitative research methods.

Through research publications and these organizations, scholars are developing and disseminating material on a variety of topics either wholly or partly focused on qualitative research. These include: case selection, counterfactual analysis, scope conditions, necessary conditions, process tracing and within-case analysis, Boolean algebra to analyze crisp or fuzzy sets, typologies and typological theorizing, concept analysis, and multi-method research strategies. The books in this review section are exemplary of the monographs published over the last few years, though there have of course been others, as well as many more articles. The articles include several significant essays in the *American Political Science Review* (APSR; see, for example, Adcock and Collier 2001; Gerring 2004; Lieberman 2005; Mahoney and Goertz 2004). These essays are important in their scholarship, but also in their venue, since publication in the "flagship" journal is one indicator of the direction the discipline is moving. According to Lee Sigelman (2005, 138, Table 3), in the 15-year period between 1985 and 2000, APSR received two submissions dealing with research methods, while in the three years between 2001 and 2004 it received 21. While some of these essays would have been on quantitative and formal methods, they included articles that are a central part of the new qualitative research methods literature.

Regardless of where one stands on the methodological spectrum, it would be hard to argue that we are not witnessing a real shift taking place in the profession. Over the last 10 years or so, the qualitative methods

research canon has been almost entirely rewritten. The new canon constitutes a renaissance in qualitative research methods that has clarified their procedures, grounded them more firmly in the contemporary philosophy of science, illuminated their comparative advantages relative to quantitative and formal methods, and expanded the repertoire of qualitative techniques (Bennett and Elman 2007c).

The author thanks Andrew Bennett, John Gerring, and James Mahoney for very helpful comments on an earlier draft. This essay draws on work co-authored with Andrew Bennett.

References

- Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Work." *American Political Science Review* 95 (3): 529–46.
- Bennett, Andrew, and Colin Elman. 2006a. "Qualitative Research: Recent Developments in Case Study Methods." *Annual Review of Political Science* 9 (1): 455–76.
- Bennett, Andrew, and Colin Elman. 2006b. "Complex Causal Relations and Case Study Methods: the Example of Path Dependence." *Political Analysis* 14 (3): 250–67.
- Bennett, Andrew, and Colin Elman. 2007a. "Qualitative Methods: The View from the Subfields." *Comparative Political Studies* 40 (2): 111–21.
- Bennett, Andrew, and Colin Elman. 2007b. "Case Study Methods in the International Relations Subfield." *Comparative Political Studies* 40 (2): 170–95.
- Bennett, Andrew, and Colin Elman. 2007c. "Case Study Methods in the Study of International Relations." In *Oxford Handbook of International Relations*, ed. Christian Reus-Smit and Duncan Snidal. New York: Oxford University Press.
- Brady, Henry E., and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman and Littlefield.
- George, Alexander L, and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Gerring, John. 2004. "What is a Case Study, and What is It Good For?" *American Political Science Review* (98): 341–54.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Lieberman, Evan. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99 (3): 435–52.
- Mahoney, James. 2007. "Qualitative Methodology and Comparative Politics." *Comparative Political Studies* 40 (2): 122–44.
- Mahoney, James, and Gary Goertz. 2004. "The Possibility Principle: Choosing Negative Cases in Comparative Research." *American Political Science Review* 98 (4): 653–69.
- Mahoney, James, and Gary Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14 (3): 227–49.
- Pierson, Paul. 2007. "The Costs of Marginalization: Qualitative Methods in the Study of American Politics." *Comparative Political Studies* 40 (2): 145–69.

Ruggie, John Gerard. 1998 "What makes the World Hang Together? Neo-Utilitarianism and the Social Constructivist Challenge. *International Organization at Fifty: Exploration and Contestation in the Study of World Politics.*" *International Organization* 52 (4): 855–885.

Sigelman, Lee. 2005. "Report of the Editor of the *American Political Science Review*, 2003–2004." *PS: Political Science* (January): 137–40.

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

doi:10.1017/S0022381607080218

In 1994, Gary King, Robert Keohane, and Sidney Verba published their *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton University Press)—hereafter DSI. This volume gathered together in a clear and accessible fashion the methodological wisdom of political science as it was then understood (at least for quantitative methodologists). Unlike traditional methods texts, DSI engaged the broad terrain of research design, with a smattering of statistical elaborations intended to elucidate broader principles. The book's subtitle indicates the authors' intent to cover "qualitative" aspects of social research. Its impact on the field has been far-reaching, though in certain respects controversial.

A decade later, a book with equal promise for transforming methodological practice has appeared. *Rethinking Social Inquiry* (hereafter, RSI), edited by Henry Brady and David Collier, is an edited volume, including contributions from the editors as well as a bevy of accomplished writers from both qualitative and quantitative traditions of research—Larry Bartels, James Mahoney, Timothy McKeown, Gerardo Munck, Charles Ragin, Ronald Rogowski, Jason Seawright, and Sidney Tarrow.¹ The product of their labors is a trenchant critique of DSI.

Yet, this is no polemic. While some readers may be expecting a clash of methodologies—DSI versus RSI, quantitative versus qualitative, positivist versus interpretivist (Harvard versus Berkeley?)—this expectation is not, by and large, fulfilled. RSI is a measured response. Perhaps the most vexing issue is the difficulty of articulating precisely what the difference between "qualitative" and "quantitative" methods might be. Chapter 13 addresses this subject

head on; however, the authors are unable to identify necessary and sufficient attributes that distinguish between these two ways of analyzing social behavior. Instead, they set forth a series of attributes that, collectively, define two ideal-types, separated by a great deal of middle ground (where, one suspects, most of the work of social science goes on). Qualitative work has a smaller N; is less likely to employ formal statistical tests; often relies on detailed knowledge of a small number of cases; and so forth (246). Other possible attributes distinguishing these two styles of work may be imagined, but none put a very fine point on the distinction, in my view. This is troubling for those who may be looking for a clear and sharp division among methodological "camps."

Thus, for those whose methodological predilections lie well outside the mainstream tradition of American social science, the debates aired in RSI may seem too restrained, too polite, altogether too agreeable. This is not just a matter of style, for there is quite a lot of common ground. Indeed, the joint authorship of both DSI and RSI includes members of both qualitative and quantitative camps; there is no rigid separation of personnel. It is not surprising, therefore, that the authors of RSI appeal frequently, and proudly, to statistical theory (e.g., to the work of statisticians like Donald Rubin and David Freedman) in the process of deconstructing DSI. The fact that one of RSI's chief purposes is to vindicate the use of qualitative methods in the social sciences should not distract us from the larger point: Brady and Collier's philosophical foundation is very similar to King, Keohane, and Verba's. This is an interne-cine battle, in other words. But this does not mean that it is unimportant. Indeed, there is quite a lot at stake.

Before turning to these issues, I want to point out several prosaic—but extremely difficult and ultimately invaluable—tasks that the book performs. First, it summarizes the lessons purveyed in DSI (in chapter two), bringing order and concision to a book that covers an expansive terrain. Second, it offers a glossary of terms—from statistics to philosophy of science—now employed routinely in both qualitative and quantitative methodological circles. This is a huge service to the field. (Even though it will not put matters of definition to rest, it does at least clarify various usages that have crept into the discourse.) Finally, it brings together a set of arguments that have circulated within methodological spheres over the past decade and articulates them in an accessible fashion for a broader audience.

¹King, Keohane, and Verba compose one chapter in the volume—a rejoinder to some of the critiques laid out by Brady et al. "RSI" refers here to all the authors except KKV.

Many of the disagreements between DSI and RSI may be understood as matters of salience, rather than valence. The authors of RSI decry the short and schematic treatment in DSI devoted to issues such as concept formation, measurement, descriptive inference, case-selection in small-N research, and the problem of explanation. Of course, DSI might respond that they have attempted a short book on a long subject and consequently had to face some hard choices about what to emphasize. Still, it is clear that the same short book would have been written quite differently had it been undertaken by Brady, Collier, and their collaborators.

One example is to be found in the way the authors define the subject matter. In addressing social science methodology DSI gives short shrift to description, an aspect of our work that is difficult to come to grips with but surely—even in DSI's estimation—crucial. And within the broad subject of causation, preference is given to techniques for estimating causal effects and the uncertainty associated with such estimations. Little mention is made of the problem of explanation, i.e., *why* might X be causing Y to occur. Proving causation is not simply about showing robust correlations; it is also, perhaps even more important, about constructing plausible explanations. Yet, the general turn toward causal mechanisms—noticeable throughout the social sciences today—is largely missing from DSI's account.

Research design sits front and center in DSI's vision of methodology. Yet, problems of methodology are viewed largely from a cross-case perspective; comparatively little is made of *within-case* methods of analysis. This is, perhaps, RSI's signal contribution and constitutes a major or minor theme of most of the chapters in this book (see, in particular, Henry Brady's appendix, which focuses on the role of "causal-process observations" in ascertaining causal relations).

Sometimes, differences in emphasis are important enough to constitute differences of opinion. Yet, one has to measure the space devoted to different subjects in order to locate these differences. This is why the DSI/RSI contrast is so subtle. It is a war of position, one might say, in which clashes on the open battlefield are rare.

This can also be seen in the area of meta-methodology, where the authors of RSI (notably McKeown) criticize DSI for neglecting a discussion of fundamentals, thereby taking a narrow "how-to" approach to the topic of methods. This is unsatisfying, even if—as seems likely—there is a good deal of agreement between DSI and RSI on what those fundamentals consist of. First, it reduces the book

to a series of assertions—thou shalt/thou shalt not. The reader is at a loss to evaluate on what basis (aside from general appeals to "science") these judgments are issued, and—where in dispute—how they might be evaluated. Second, DSI's rulebook approach conceals the manifold ways in which our philosophical understandings about science, and our ontological understandings about the real world, may influence our study of social behavior. Finally, DSI's nose-to-the-ground approach by-passes the influence that philosophical currents have had on current understandings of methodological practice. I am thinking here, in particular, of the rise of "realism" and the coincident fall of "positivism" in the literature emanating from philosophy of science over the past several decades. No doubt, many social scientists receive a watered-down version of these debates; still, their influence may be considerable, and appears to be growing, judging by the number of citations that now appear on journals and books from the political science community.

A key critique of DSI is that the book lays out a template for social science endeavor that, if taken seriously and followed strictly, would exclude a great deal of extant work in the field, including a majority of work in the qualitative *and* quantitative traditions. Writers point out that certain key criteria identified as critical to causal inference by DSI (borrowing from work by Ruben and Holland) such as causal homogeneity and conditional independence are unmet—or are met only with the most dubious assumptions—in work based on observational data, whether small- or large-N. In this respect, DSI is setting forth a standard that is purist to the point of perfection. DSI respond that they are trying to improve the practice of social science, not to write out certain sorts of endeavor—*better* qualitative and observational work, in other words. However, RSI points out that many of the injunctions in DSI are very difficult, if not impossible, to implement in an observational setting. So, one is faced with a conundrum of abandoning this sort of work, or settling for a much lower level of certainty in one's results.

The consequences of the DSI model for qualitative research seem particularly severe. To be sure, small-N projects *could* be converted into large-N projects and thereby go some way toward fulfilling the numerous and stringent desiderata specified in DSI. But the cost of doing so might be—and presumably often would be—to change the very nature of that research. For example, the result of such a conversion from science (qua RSI) to science (qua DSI) might involve a sacrifice of insight for the sake

of gains in certainty. We are all familiar with the solid inference (“statistically significant at 95%”) that adds little to our store of useful knowledge.

This leads to a final point. DSI is accused of issuing rules and ignoring the hard fact that such rules of inference often conflict with one another. From RSI’s perspective, designing successful research is better understood as a matter of dealing with tradeoffs across multiple criteria of adequacy rather than following necessary and sufficient rules. One such tradeoff, underlying the entire subject of methodology, is the pull between “diverse tools” and “shared standards”—the paired terms contained in RSI’s subtitle. The rub is that tools contain within them methodological strengths and weaknesses; hence, the choice to employ them involves an implicit choice among methodological criteria. For example, experimental and case study methods tend to privilege internal validity over external validity, while large-N observational methods carry the opposite set of strengths and weaknesses. There is no obvious way to resolve such differences. While we can all agree that both internal and external validity are important, we shall probably never agree on their prioritization. In this fashion, the diversity of methodological tools is forever undermining a consensus over standards. To put the matter another way, if all social scientists shared precisely the same standards we would also use almost precisely the same set of tools. And if there were no consensus whatsoever on standards, there would be no theoretical limit on the sort of tools that might be applied to social science problems, and no way to adjudicate amongst them.

Some readers will view RSI as a “corrective” to DSI—addressing issues that DSI neglected, clarifying points that remained obscure, and complicating issues that are dealt with in a schematic fashion. This is consistent with the approach taken by Brady et al., who have elected to write a rejoinder to DSI, rather than provide a fresh account of the subject. Others will see RSI as elucidating, or at least defending, a very different way of doing business, one that does not fit into the “quantitative template.”

Before concluding, it should be pointed out that many of the omissions charged to DSI are not easily addressed. Can one say something meaningful—i.e., moving beyond common sense—about how to enlist qualitative evidence (e.g., causal-process observations), how to formulate and test descriptive inferences, how to conduct ethnographic research, how to explain correlations, how to formulate useful hypotheses and connect them with larger theories about the world? Arguably, DSI has taken the easy portion of

social science methodology—how to reach conclusions about a causal effect once the data are in—and left the rest for posterity to figure out. In this respect, RSI constitutes a strong challenge to a future generation of methodologists. By raising these issues, and pointing to the deficiencies of the quantitative template, an ambitious agenda has been laid on the table.

John Gerring, *Boston University*

Case Studies and Theory Development in the Social Sciences. By Alexander George and Andrew Bennett. (MIT Press, 2005.)

doi:10.1017/S0022381607080231

Alexander George and Andrew Bennett have given us a book of potentially very broad usefulness. Not exclusively a work of methodology, *Case Studies and Theory Development in the Social Sciences* discusses issues in the philosophy of science, practical challenges involved in research design and execution, the state of the debate on democratic peace, and theory and method for within-case analysis. The presentation of each of these themes is richly enhanced by often detailed examples from substantive research; an appendix adds to the embarrassment of riches by providing an additional 39 pages of commentary on various published research projects.

First and foremost, this volume is a compendium of pragmatic wisdom about the often challenging process of carrying out research on political science topics. At the heart of the book are four crisply written chapters on “How to Do Case Studies” (Part II of the volume, comprising Chapters 3 through 6). While the advice in these chapters is framed in terms of case-study research, the large majority of this material may be valuable to analysts, and especially graduate students, designing research based on any empirical method. Qualitative and quantitative researchers alike face the challenges of: finding ways for their research to contribute to theory; defining a problem and a research objective; characterizing the primary unit of analysis; choosing specific independent and dependent variables; selecting cases; characterizing the range of observed differences on key variables; collecting data; dealing with theoretically relevant surprises during the research process; assessing the quality of measurements; and determining the proper contextual bounds for one’s conclusions. George and Bennett offer incisive advice and practical wisdom regarding each of these topics.

One example is the book's discussion of the challenges of reconstructing the process behind policymaking decisions (94–99). In this context, the authors warn that political scientists “should forgo the temptation to rely on a single, seemingly authoritative study of the case at hand by a historian” (95), instead acknowledging the diversity of interpretations in most historiography and translating each perspective into a hypothesis to be tested during the analysis. Obviously case-study scholars, or qualitative analysts, have no monopoly on the use of published history, either to set the stage for their research or to rule out certain hypotheses as inconsistent with what is known about a given place and time. George and Bennett's caution to treat the historical literature as a debate rather than a source of consensus would thus seem broadly relevant.

In the same section, the authors provide more widely applicable practical wisdom in their warning against “over-intellectualizing the policy process” (98–99). Here, the researcher is warned against theories of the policy decision-making process that assume an excess of order and rationality, but the special focus is on warning against the temptation to offer univariate accounts of political decisions: “presidents and top-level executives often seek multiple pay-offs from any decision that they take” (98). For instance, a policy may be selected in part because it is expected to have good macroeconomic consequences, but also because it will be politically popular and will allow patronage goods to be distributed to key constituencies. A theory that focuses on only one of these factors will necessarily be incomplete and may therefore fail to adequately explain other decisions in which all of these factors are not aligned. This sensible warning clearly applies in political science domains far beyond case studies of the policymaking process. Presumably, decision makers are generally free to draw on multiple considerations, and theories of decision making in electoral, social-movement, or military contexts, for example, may require just as much attention to multivariate causation as theories of policymaking.

These are merely two examples of the widely applicable practical wisdom about the research process that permeates *Case Studies and Theory Development in the Social Sciences*. Such advice makes this book, and indeed much of the literature on qualitative methods, of very real value to empirical researchers, especially those with little experience, regardless of the methods they plan to employ.

Other parts of the volume are more narrowly tailored to the task of defining, specifying the limits

of, and refining strategies for specific qualitative case-study methods. George and Bennett concentrate in particular on a method of within-case analysis that they refer to as “process tracing,” a term that George imported from psychology into qualitative methods some time ago. The authors define process tracing as a method in which “the researcher examines histories, archival documents, interview transcripts, and other sources to see whether the causal process a theory hypothesizes or implies in a case is in fact evident in the sequence and values of the intervening variables in that case” (6). George and Bennett argue that there is a special congruence between this empirical approach and those strands of the philosophy of science that emphasize the discovery of causal mechanisms. Their argument, presented especially in Chapters 7 and 10, is intricate and represents a real advance in specificity and rigor for the literature on the justification for qualitative within-case analysis.

Even so, some ideas central to the connection between qualitative within-case evidence and causal mechanisms are passed by without comment. Causal mechanisms are a component of theory; as George and Bennett note, “Theories or models of causal mechanisms must undergird each step of a hypothesized causal process...” (147). Because we are operating in the realm of theory, empirical information of any sort will require inference in order to connect with causal mechanisms. Yet George and Bennett never address the inferential nature of within-case research on causal mechanisms, failing to discuss the tasks of conceptualization, measurement, and the making of assumptions regarding unanalyzed potential explanations. This is important as a matter of method; qualitative within-case research could surely benefit from more frequent systematic discussion of such matters. It is also important in terms of comparing the fit between various methods and the goal of understanding causal mechanisms. Other techniques, after all, also permit inference about causal mechanisms. Laboratory experiments, for example, may allow tests of hypotheses related to causal mechanisms at the individual level, even for research questions connected with macrolevel phenomena. Clearly, experiments and process tracing have divergent strengths and weaknesses for inference regarding causal mechanisms. An exploration of the trade-offs involved in choosing one of these methods would move the discussion much farther than an incomplete assertion of congruence between causal mechanisms and within-case qualitative analysis.

The book misses other opportunities of this methodologically comparative kind. Particularly, in their various chapters related to process tracing, the authors never mention the influential and widely practiced laboratory and survey-research versions of process tracing. These techniques include specific tools such as computerized information boards, tracking eyeball movements, and instructions to experimental subjects or survey respondents to talk aloud through their thought process while answering a question or making a decision. A careful exploration of parallels and divergences between these forms of process tracing drawn largely from psychology and the more historically grounded versions of process tracing that George and Bennett emphasize has the potential to bring qualitative within-case analysis more centrally into social science methodological discourse. Such missed opportunities point to directions for further work in the growing subfields of research on qualitative and mixed methods.

The book's overall value, largely as a source for essential practical wisdom about social science research but also as a contribution to the methodology of qualitative within-case analysis, far outweighs the occasional distraction resulting from topics that are perhaps overemphasized or discussed in potentially misleading ways. One such topic is equifinality, which the authors define as the existence of "many alternative causal paths to the same outcome" (10). The book routinely treats equifinality as a form of causation for which quantitative analysis is not especially adequate and for which case studies are particularly well-suited (e.g., 9–10, 19–20, 215). But in fact, the potential for discovering equifinality is built into most additive statistical models. In a logit model with two dichotomous independent variables, for example, the result that both variables have very large, positive, and statistically significant coefficients would effectively be a finding of equifinality: the outcome has a quite high probability of occurring if either of the independent variables has a score of 1. So equifinality raises few distinctive issues for quantitative research. By contrast, as George and Bennett point out, equifinality severely limits the validity of case-study research using Mill's methods or related elimination techniques (157–58). Hence, equifinality would in fact seem to be an example of a problem which distinctively challenges case studies, not quantitative methods.

This and other similar minor distractions in no way undermine the very real value of this book for political scientists, especially graduate students and other beginning empirical researchers, regardless of

their methodological predilections. We all face issues of theory building, conceptualization, measurement, and building evidence of causation. The treasure trove of wisdom, practical advice, and methodological commentary that George and Bennett provide should improve the ability of most readers to successfully address these shared dilemmas.

Jason Seawright, *Northwestern University*

Comparative Historical Analysis in the Social Sciences.

Edited by James Mahoney and Dietrich Rueschemeyer. (Cambridge University Press, 2003.)

doi:10.1017/S0022381607080279

This is an excellent book that includes thoughtful and readable contributions by some of the leading methodologists and practitioners in the area of comparative historical analysis: Edwin Amenta, Jack Goldstone, Roger Gould, Peter Hall, Ira Katznelson, James Mahoney, Paul Pierson, Dietrich Rueschemeyer, Theda Skocpol, and Kathleen Thelen. I fully share the editors' view that comparative historical analysis is and should be a "leading mode of analysis, widely used throughout the social sciences" (5). The individual chapters and the book as a whole will be of considerable value to anyone interested in better understanding it.

A large number of important questions are addressed in the volume. For me the most interesting have to do with the analytical utility of what is often referred to as "small-N comparison" but what I will call "small-N analysis": To what extent is this type of analysis useful for theory testing? And what role does cross-case (usually cross-country) comparison play? For Mahoney and Rueschemeyer (10–15), a concern with causal analysis (as opposed to description) and cross-case comparison are two of the three defining features of comparative historical analysis (the other is a concern with over-time processes).

My conclusion is that where small-N analysis is most effective at contributing to explanation (causal analysis), cross-case comparison plays no role. Cross-case comparison in small-N analysis is useful mainly for theory building rather than theory testing.

Mahoney's chapter on "Strategies of Causal Assessment in Comparative Historical Analysis" nicely details the principal ways in which small-N analysis can contribute to hypothesis testing. One is what he calls "ordinal comparison." Here cross-case comparison is critical. The analyst performs what is essentially a correlational analysis—rank ordering or scoring the cases on one or more independent

variables and an outcome variable and seeing whether they correlate strongly and in the expected direction. I suspect this is what most researchers initially have in mind when they decide to do a small-N comparative analysis.

This analytical strategy has two significant and well-known limitations. One is the possibility of omitted variable bias. With only a few cases, there may be many potentially relevant causal factors on which the cases are similar. It is thus difficult to feel very confident about the relative importance of any particular one or two or three of them. This leads many analysts to pursue a “most similar cases” design. If the countries share many features but differ on the one(s) of interest to the analyst, omitted variable bias is less likely. For example, the United States might be compared to Canada, or Sweden to Denmark. This is a wise strategy, yet is hardly foolproof; many country experts would be appalled by an assertion that Canada is like the United States or Denmark is like Sweden “in most relevant respects.”

The second limitation of ordinal comparison is generalization. Proponents of small-N analysis sometimes argue that their aim really is only to understand the particular cases they are studying. But I doubt this is the objective of most, and a number of the contributors to this book say explicitly that generalization should be a goal. Even if an ordinal comparison yields a strong correlation and there is reason to think there is no omitted variable bias, an N of two or three or four must raise concern about whether the finding applies to other cases.

Because of these limitations, ordinal comparison should be viewed mainly as an exercise in theory building rather than theory testing. (I think much “large-N” quantitative analysis should also be viewed as theory building, but that is an issue for a separate discussion.) Or it can be *part* of a cumulative theory-testing enterprise that consists, as Goldstone suggests in his chapter, of multiple small-N studies of the same hypothesis(es) across, eventually, a larger number of cases.

A second way Mahoney suggests small-N analysis can contribute to theory testing is via elimination of a hypothesized necessary or sufficient condition. If the theory holds that the particular causal factor is always (as opposed to usually or nearly always) necessary or sufficient, a single inconsistent case is enough to invalidate the hypothesis. But here comparison across the cases is of no analytical relevance.

A third is process tracing (also discussed in the chapters by Hall, Goldstone, and Amenta), in which the analyst looks closely at the role of various hypothesized causal factors in one or more cases to

see if they appear to have actual causal relevance—in other words, to see if there are perhaps other causal factors that render the correlation spurious. Here too cross-case comparison is not critical. The key is to focus on the details of each particular case.

A fourth is examination of multiple observations within one or more cases. The researcher may conduct what is in effect a time-series analysis for a single country or analyze across subnational geopolitical units (also discussed in Rueschemeyer’s chapter). Once again, though, comparison across the cases (nations) plays no role in causal assessment.

Thus, cross-case comparison is useful in small-N analysis mainly when the analytical approach is that of ordinal comparison, but ordinal comparison is very limited for purposes of theory testing. In the most useful small-N theory-testing approaches, cross-case comparison plays no role.

This point has important implications for case selection for those interested in conducting a small-N analysis. If the principal analytical aim is process tracing or testing a hypothesized (always) necessary or sufficient condition, there is no need to worry about whether the cases can be usefully compared. If the aim is theory testing via comparison across the cases, the comparison does need to be justified, and the analyst should make it clear whether she or he wants to draw inferences about only those particular cases (and if so, why) or whether the analysis is intended to be part of a larger cumulative endeavor that will eventually include other cases.

Lane Kenworthy, *University of Arizona*

Necessary Conditions: Theory, Methodology, and Applications. Edited by Gary Goertz and Harvey Starr. (Rowman and Littlefield, 2003.)

doi:10.1017/S0022381607080255

The editors and contributors of this book cogently demonstrate how the fairly simple looking and often misunderstood concept of necessary conditions can generate various intriguing methodological and theoretical consequences for social science research—regardless of the method used and the research topic studied. This book engages everybody in critical self-reflection by showing that—contrary to widespread belief—necessary conditions are not rare in social science theorizing (Goertz in Chapter 4 alone counts more than 150 in the social science literature of the last three decades); they do not imply a deterministic notion of causality which by some is deemed alien to modern social sciences; they do not inevitably require

the use of dichotomous data; they do not have to be spurious or trivial; and they are currently not adequately dealt with in research based on standard statistical practices, primarily because the notion of correlation—the cornerstone of most quantitative statistical analyses—is inappropriate for investigating statements of necessity.

This book deserves special credit for not just opening a Pandora's Box but giving hints at how to close it by outlining novel statistical procedures designed for analyzing necessary conditions that incorporate the rich and decades-old body of literature on probability theory and statistical tests. This anthology should appeal to scholars from all methodological schools and puts pressure on us all, for after this book nobody can claim not to know about the theoretical importance and methodological intricacies of necessary conditions and how to methodologically tackle that complexity.

Almost in passing, this book offers thought-provoking insights in universal research design issues such as case selection, scope conditions, and model specification. For instance, various authors (notably Most and Starr in Chapter 2) convincingly show that, contrary to standard advice to graduate students, we not only should select cases on the dependent variable, but also ensure there is little or no variation on it. If necessary condition hypotheses are tested, such case selection is a logical must.

By thoroughly spelling out the different aspects of the logic of necessary conditions, this book shows that, currently, on the one hand, some scholars say necessary condition but don't mean it, while on the other hand, many people mean necessary conditions but don't say it. That is bad for social science research because it means we are likely to test our theories with inappropriate methods.

The 13 chapters are a refreshing mix of abstract methodological reflections, hands-on suggestions, and real-life research applications. The introductory chapter by Goertz and Starr does a marvelous job of systematizing the debate on the wide array of issues discussed in this book. It helps to better understand how the different contributions covering very different topics and methods speak to each other and why they are relevant for understanding necessary condition theory, methodology, and application.

Based on their carefully developed argument that there are "various interpretations of the necessary condition concept" (see especially page 12; Table 1.1), Goertz and Starr pave the way for a methodological pluralism in dealing with necessary conditions in the social sciences. Most importantly for quantitative

scholars, the data can be continuous rather than just dichotomous and the hypothesized necessary condition relationship can be of probabilistic rather than just deterministic nature. None of these concepts can claim to be the true concept of necessary conditions. While interlinked, all highlight different aspects and all are based in different schools of thought: mathematical and philosophical logic, (fuzzy) set theory, calculus, and probability theory. The message that there is more than one way of thinking about necessary conditions is not to be confused with an "anything goes" attitude. Rather, this book shows that each conception of necessary conditions comes with a set of specific rules how to correctly test them.

Levy's chapter on the process leading to the outbreak of WWI exemplifies the claim that case studies show a strong affinity to the language of necessary conditions. What could have been highlighted more is that the plausibility of necessary conditions in case studies rests upon the plausibility of the counterfactual arguments put forward. This reliance on counterfactuals sets case study approaches to necessity apart from large(r) N tests of necessary conditions, which are based on empirical distributions of cases, as demonstrated in the chapters by Braumoeller and Goertz, Dion, Ragin, and Tsebelis.

Not all cases matter in statistical tests of necessity, though, and including them—as most common measures of association, such as chi-square, gamma, or tau-beta, do—generates flawed results. This is probably the book's most forceful and consequential deconstruction of "standard statistical reflexes" (197) when approaching necessary condition hypotheses. For example, the hypothesis "rich cases are democratic" might sound similar but is utterly different from "the richer a case the more democratic it is." The first postulates an asymmetric set relation (democraticness is claimed to be necessary for richness, or richness sufficient for democraticness, respectively) and the second a symmetric correlation between the variables "richness" and "democraticness." The necessity (or sufficiency) statement would be confirmed if a triangular heteroskedastic pattern was found in the data whereas the correlational statement expects us to find homoskedasticity with most cases on or close to the regression line. In short: "necessary condition does not equal correlation" (48). Unsurprisingly, this implies severe limitations to dealing with necessary conditions within the framework of correlation-based standard statistical techniques. Tsebelis (Chapter 11) makes an innovative attempt at combining familiar statistical tools, but truly adequate statistical procedures for dealing with necessity are

probably more complex and break more radically with common statistical practices. This is the message one gets from reading Braumoeller's and Goertz' ideas in Chapter 9.

The above example showing that set relations and covariations are different raises an important question that remains somewhat unresolved in the book. How should one overcome the tension between the verbal formulation of a necessary condition hypotheses framed in terms of sets, or, types of cases (e.g., "rich" vs. "not rich"), on the one hand, and empirical tests based on continuous data? By definition, with a continuous measure there is no clearly specified level at which cases are "rich" or "not rich." This is problematic because all of the most thought-provoking arguments in this book—select on the DV and have no variation, do not use information in the 0,0 cell of a 2x2 table, do not use standard measures of associations—rest on the notion that the scales for measuring the condition and the outcome have a starting and an endpoint with distinct qualitative meanings. This is not a plea for limiting necessary condition hypothesis testing to (crisp or fuzzy) set theoretic approaches, but more research needs to be done on how these well-grounded criticisms of today's research approaches to testing verbally formulated necessary condition hypotheses translate once necessity (and sufficiency) are analyzed with continuous data void of any set-theoretic meaning.

A book opening up and partially closing so many fundamental research methodological issues should probably not be criticized for what has not been addressed. What is more, since its publication some of the issues left open have by now been covered (with active contributions by some of the editors and contributors of the book). Examples include coefficients expressing the empirical consistency and relevance of necessary (and also sufficient) conditions by Goertz (*Studies in Comparative International Development* 41 [2]: 88–109) and Ragin (*Political Analysis* 14 [3]: 291–310). In addition, scholars now have available Ragin, Drass and Davey's (2006) software for dealing with necessary (and also sufficient) conditions based on data expressing set memberships, as well as Ragin's (2007) work on advancements in set membership based on raw data (in *Oxford Handbooks of Political Science*). Finally, Goertz and Levy have developed systematic treatments of necessary conditions not just in causal inference but also in concept formation in Goertz's book, *Social Science Concepts* (2006), and specifications of the role of counterfactuals for

assessing necessary conditions in case studies in Goertz and Levy's (2007) book *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*.

Thus, the following reads more like a wish-list of topics the authors hopefully will write their next book(s) on: Sufficient conditions are mentioned throughout the book but their implications for social science theory and practice are not dealt with systematically. The prevailing standpoint in this book is that, mathematically and logically speaking, necessity can easily be transformed into sufficiency. Such a prevalence of pure formal logical laws over social scientific reasoning is not fully satisfying. It makes a big difference for theoretical and research design choices whether one analyzes necessary conditions for, say, democracy, as opposed to sufficient conditions for nondemocracy (Harvey in Chapter 7 demonstrates this forcefully). Furthermore, as Cioffi-Revilla in his very nice concluding chapter points out, paying attention to both necessary and sufficient conditions is quite natural and ultimately more insightful (297). There is, however, also no doubt that bringing sufficiency on board adds yet another level of complexity to social science theory and methodology, as one has to simultaneously cope with issues of equifinality (different conditions lead to same outcome), multifinality (same condition leads to different outcomes), conjunctural causation (combinations of conditions lead to outcome), asymmetric relations (occurrence and non-occurrence of outcome require separate analyses and explanations), and INUS conditions. Braumoeller makes some valuable suggestions on how to deal with all these issues within a statistical framework but more needs to be written on how to handle that complexity within the quantitative—but also the qualitative—research template, especially if the so far fairly neglected time dimension (plus timing and sequencing) is added to the notion of causal complexity in terms of necessity and sufficiency (see Braumoeller, *Political Analysis* 11 [3]: 209–33).

All contributors to this book deserve praise for presenting difficult methodological issues in a clear and understandable way. Such clarity is a necessary condition for further progress in this important area of research. By doing such a wonderful job in terms of clarity and readability, the editors certainly prove themselves wrong with their hunch that their book, by showing the intricacies of necessary condition hypotheses, might lead to "an increased hesitance in proclaiming necessary condition hypotheses" (22). Instead, it stimulates interest in necessary conditions and definitely helps to get tests of necessity right, no

matter in which research tradition these tests are grounded.

Carsten Q. Schneider, *Central European University*

Case Study Research: Principles and Practices. John Gerring (Cambridge University Press, 2007).
doi:10.1017/S0022381607080243

Case-study research, as John Gerring points out in his erudite primer on the method, occupies a central position in a great variety of social science disciplines. This suggests a paradox, however, because case studies are often viewed with circumspection, even in disciplines responsible for a large output of actual case-study work. Along with several other recent contributions in political science, Gerring's new book seeks to put case studies on firmer methodological ground, illuminating the inferential logic and value of the case-study research design.

The book is also concerned with the relationship between case studies and other methods. Gerring suggests, provocatively, "there is no such thing as a case study, *tout court*. To conduct a case study implies that one has also conducted cross-case analysis, or at least thought about a broader set of cases (13) . . . Case study analysis does not exist, and is impossible to conceptualize, in isolation from cross-case analysis" (90). The point is well-taken, for only from this perspective does the familiar question asked about case studies—what is this a case of?—make sense. Yet there are other issues to explore. How should cross-case analysis inform the selection and implementation of case studies? How does the inferential logic of the case study relate to the logic of other methods? What are the distinctive sources of inferential leverage provided by case studies relative to other methods?

Gerring offers several kinds of answers. In a chapter coauthored with Jason Seawright, the authors elaborate a range of strategies for using cross-case (often regression) analysis to select cases for intensive study. These include strategies familiar from other writings on case-study research, such as the selection of crucial (Eckstein in *Handbook of Political Science*, 1975) as well as most-similar and most-different cases (J.S. Mill, *A System of Logic*, 1843; Przeworski and Teune, *The Logic of Comparative Social Inquiry*, 1970). Gerring and Seawright also advocate the use of various regression diagnostics to identify "typical," "deviant," "influential" and other kinds of cases. They describe increasingly popular statistical matching procedures as a useful way to formalize a notion of most-similar cases.

The authors concede that the regression techniques they advocate are heavily model dependent; if the regression model that is fit to a large N dataset is misspecified, then there may be little value in using the model to choose cases for intensive study. Thus, while the chapter by Gerring and Seawright offers advice on how to merge quantitative and qualitative methods, the cautionary notes may be even more important than the advice.

In another chapter, coauthored with Rose McDermott, the logic of case studies is compared to the logic of experiments. Following Jerzy Neyman ([1923] 1990), much recent writing on randomized experiments emphasizes the following inferential problem. Imagine a medical patient who can either take a pill or not take the pill; the patient cannot do both. In the experimental template, the *causal effect* of the pill on the patient's health is the difference between what would happen to the patient with the pill and without it. The causal effect is therefore unknowable. A randomized controlled experiment solves the inferential problem by estimating, not the causal effect of the pill for a particular patient, but rather the average causal effect for all patients in some defined universe. Gerring and McDermott lay out the logic of "counterfactual comparison" underlying experiments and advocate its usefulness for thinking about causal inference in case studies (165–68).

Yet the authors' claim that the logic of experimental inference contains broader lessons for case studies, beyond providing a sharp way to define causal effects, seems exaggerated. In actual experiments, causal effects are estimated by assigning some of the patients at random to receive the pill and others at random to control; estimates of causal effects get more precise as the groups randomly assigned to treatment and control conditions grow in size. To get around the problem of estimating the causal effect of treatment for a particular patient by estimating the average causal effect for all patients in the universe, however, it is necessary to have at least a medium-sized N. With only a few cases at hand, the experimental template may have little relevance, except as a way of defining causal effects—when causation is to be understood in terms of interventions. Nonmanipulationist accounts of causation are also little discussed in this volume, yet they may play an important role in much social science research (Goldthorpe, *European Sociological Review* 17 [1]: 1–20).

One reason that Gerring finds a discussion of experimental and related quasi-experimental templates useful, I believe, is that he does not understand case-study research as an inherently small-N method.

According to Gerring, “each case may provide a single observation or multiple (within-case) observations (19) . . . For those familiar with the rectangular form of a dataset [i.e., a data matrix], it may be helpful to conceptualize observations as rows, variables as columns, and cases as either groups of observations or individual observations” (22). Gerring then advises the use of experimental or quasi-experimental methods to analyze variation across within-case observations. Yet for practitioners of case studies, the advice seems problematic; as more and more observations are compared, and the inferential leverage afforded by experimental or quasi-experimental methods is brought to bear, it seems less likely that one is really conducting a case study—that is, an intensive analysis of one or several instances of a phenomenon (19–20).

The concluding chapter on process tracing, co-authored with Craig Thomas, seems to come closest to shedding light on the distinctive contributions of case-study research to causal inference. In process tracing, not-strictly comparable pieces of information are combined in a way that adds up to a convincing causal account, by rendering alternative explanations less plausible while showing that microevidence is consistent with theoretical claims. Process tracing may be akin to detective work; bits of evidence about the maid, the butler, and the suspect are combined to formulate or investigate a central hypothesis about who committed a crime. In a different but related account, Collier, Brady, and Seawright (in *Rethinking Social Inquiry*, 2004) describe how what they term causal-process observations can provide a smoking gun that demonstrates—or rules out—a particular causal hypothesis. Unfortunately, the discussion of process tracing is a relatively short addendum to the core concerns of Gerring’s book.

Definitional slippages are one distracting feature of this book, despite the inclusion of a glossary and careful attention to defining terms. At one point, for instance, Gerring notes parenthetically, “I use the terms proposition, hypothesis, inference, and argument interchangeably” (22). Nevertheless, articulating the distinctive contributions to social science of case studies has been a core challenge for qualitative methodologists, and together with other recent contributions in this area (Brady and Collier, *Rethinking Social Inquiry*, 2004; George and Bennett, *Case Studies and Theory Development*, 2005) Gerring’s volume takes on this challenge in a spirited fashion. It will provide useful and engaging reading for substantive researchers of all methodological stripes.

Thad Dunning, *Yale University*

Politics in Time: History, Institutions, and Social Analysis. By Paul Pierson. (Princeton University Press, 2004.)

doi:10.1017/S0022381607080280

In *Politics in Time*, Paul Pierson has written an important book that is engaging, ambitious, and provocative. Its purpose is essentially threefold: to advocate that political scientists situate arguments in temporal perspective, to illustrate a number of ways in which they might do so, and to argue that much of the discipline does not presently take time seriously enough. The book is oriented not merely toward qualitative scholars or historical institutionalists, but rather “those interested in the attempt to develop claims about the social world that can potentially reach across space and time” (7)—effectively all social scientists who seek to advance generalizable explanations. In his effort to reach such a broad audience, Pierson engages widely with the discipline and beyond, drawing upon theoretical insights from Kenneth Arrow to Arthur Stinchcombe, exploring causal mechanisms from positive feedback to absorbing Markov chains, and offering substantive examples from U.S. congressional committees to state building in early modern Europe.

The publication of Pierson’s book in 2004 occurred in the midst of a still-ongoing resurgence of interest in qualitative methods and temporal arguments—stimulated in no small part by Pierson’s *American Political Science Review* article on path dependence published in 2000. Following an initial series of articles on the sources of institutional lock-in, a second wave of scholarship by authors such as Kathleen Thelen and Jacob Hacker shifted the focus to ways in which institutions change rather than remain stable over time. *Politics in Time*, which presents revised versions of four previously published articles as well as an entirely new introduction, fifth chapter, and conclusion, plays the very useful role of encapsulating this evolution of the literature and also offering an attempt at synthesis. Ultimately, as Pierson argues in Chapter 5, change and continuity must be seen as two sides of the same coin.

Politics in Time begins with a focus on institutional continuity via path dependence and positive feedback. Examining the mechanisms that sustain stability over time in economic history, Pierson argues that such processes should be at least as common in the political realm. Reasons include the prevalence of collective action in politics; the potentially self-reinforcing accumulation of power asymmetries; the absence of a price mechanism to clearly indicate

optimal behavior; and the centrality of formal institutions, which have large set-up costs and generate learning effects, coordination effects, and adaptive expectations. When political actors begin to travel down a self-reinforcing path—even one that is eventually considered inferior to the road not taken—reversal is unlikely because of the short time horizons of many political actors and the obstacles to change built into many political institutions.

Having established the importance of positive feedback mechanisms, the second chapter of Pierson's book examines the closely related issue of timing and sequence. Here, he advocates an approach that combines the insights and precision of rational choice models of legislative cycling with the attention to large-scale social changes that have been the mainstay of historical institutionalism and American political development. Sequencing arguments, he maintains, need not be "Dr. Seuss-style," ad hoc explanations in which it "just so happened" that one event followed another. Rather, the best of these arguments illustrate, often via comparison, how the order of particular events matters crucially for the eventual outcome. Examples include the self-reinforcing incumbency advantages enjoyed by a party that initially "fills up" a political space, such as by organizing the working class. Future competitors for this constituency will have greater difficulty than if they had been the first ones to occupy the same political niche.

The third chapter of Pierson's book shifts the focus from arguments emphasizing continuity to those that focus on long-term change. Many topics in political science, Pierson argues, are those in which cause immediately precedes effect and both occur relatively quickly. The tendency to limit time horizons of research topics in this fashion, however, ignores a number of important processes, such as cumulative causes, threshold effects, and causal chains, in which the cause or the outcome unfolds over a significant period of time or there is substantial temporal separation between the two. Pierson argues that the trend toward research involving only short-term independent and dependent variables risks excluding many of the classic subjects in political science, such as party-system change and state building. An exclusively short-term focus may even contribute to errors of causal inference if scholars mistakenly treat slow-moving causal variables as mere background conditions.

An excessively static approach to political science can occur not only in the selection of research topics, but also in the explanation of institutional origins—an

argument Pierson develops in the fourth chapter. Here, *Politics in Time* takes aim at "actor-centered functionalism"—the rational-choice claim that institutions exist because farsighted, purposive, and instrumental actors benefit from them. Drawing upon numerous empirical examples, Pierson lays out six limitations to this perspective. Institutions may have multiple or unanticipated effects, such that their existence cannot be explained by a simple reading of their creators' preferences. Designers also may adopt institutional forms for noninstrumental reasons, such as diffusion or cultural specificity. Actors may have short time horizons, creating institutions with undesirable long-term effects, or their preferences may change over time as institutions remain stable. Finally, political actors themselves may change; a new generation may inherit institutions that reflect previous actors' preferences rather than their own.

In recognition of these limitations to the rational choice perspective, many historical and sociological institutionalists have begun to articulate mechanisms of how institutions change over time. Yet in Chapter 5, Pierson argues that shifting from an assumption of static institutions to an exclusive focus on fluidity involves too much of a swing of the pendulum. Research on institutional change, which has identified mechanisms such as critical junctures, layering, conversion, and diffusion, places undue emphasis on the malleability of institutions at the hands of entrepreneurs or the losers from previous rounds of political competition, and it does not predict when we should expect one type of institutional change versus another. Most importantly, arguments about institutional change often lose sight of the sources of institutional stability, such as coordination problems, veto points, asset specificity, and positive feedback. Thus, Pierson concludes by proposing five distinct research agendas on processes of institutional development that seek a synthesis between arguments about change and continuity.

As mentioned at the outset, the objective of Pierson's book is not only to illustrate ways in which political scientists can incorporate the temporal dimension into their research, but also to argue that much of the discipline does not presently take time seriously enough. While acknowledging that rational choice scholars and quantitative methodologists often *can* model long-term historical processes, Pierson insists that theoretical possibility does not translate into typical practice. The main evidence to support this claim is presented in Chapter 3, where he shows that in several top political science journals, articles investigating short-term causes and outcomes vastly

outnumber those in which change in an independent or dependent variable occurs over a long period of time. Yet in attributing this outcome to the rise of rational choice analysis and quantitative methods, Pierson moves into the realm of assertions that—while plausible—he does not have data to support.

Similarly, Pierson argues that statistical tools for estimating complex temporal relationships “frequently go unexploited” (168), and that while it is possible to incorporate long-term sociological processes into rational choice explanations of comparative politics, “they are not the kinds of hypotheses that these analysts typically go looking for” (100). Yet it is questionable whether typical practice is so myopic at a time when Beck and Katz’s 1995 article on time-series cross-sectional methods is the ninth most cited article in the history of the *American Political Science Review*, and Acemoglu and Robinson’s (*Economic Origins of Dictatorship and Democracy*, 2005) rational choice analysis of long-term processes of democratization has generated so much excitement among comparativist formal modelers. Quantitative and rational choice scholars certainly approach the temporal dimension in a different manner than does Pierson, but his point is that it is rare for these scholars to take time seriously *at all*. His critiques of tendencies in the discipline may ultimately be accurate, yet they would be much stronger if they were accompanied by a more convincing empirical demonstration of actual practice.

Despite overstepping the bounds of its own evidence when critiquing current practice on the quantitative and formal side of the discipline, Pierson’s book nonetheless offers an excellent demonstration of best practice on the qualitative side as well as a convincing claim that all political scientists should take time seriously. Some will certainly dispute the book’s arguments, but one suspects that that is precisely the author’s intention. Indeed, the greatest value of *Politics in Time* lies in the fact that is certain to stimulate (and indeed, has already begun to stimulate) an important process of debate about temporal processes in politics and the ways in which different research traditions can better address them in the future.

Taylor C. Boas, *University of California, Berkeley*

Progress in International Relations Theory: Appraising the Field. Edited by Colin Elman and Miriam Fendius Elman. (MIT Press, 2003.)

doi:10.1017/S002238160708022X

Colin Elman and Miriam Fendius Elman have produced a refreshingly coherent, thought-provoking, and engaging volume dedicated to the use of Imre Lakatos’s methodology of scientific research programs (MSRP) for appraising progress in the field of International Relations (IR). After the editors’ detailed overview of MSRP and the controversies it has generated in the philosophy of science community and other disciplines, Part I of the book moves into a series of applications by leading figures in IR. Robert Keohane and Lisa Martin reconstruct institutional theory according to MSRP, followed by Jonathan DiCicco and Jack Levy on power transition theory, Andrew Moravcsik on liberal IR theory, James Ray on the democratic peace, Stephen Walker on operational code analysis, Robert Jervis on the debate between realist and neoliberal IR theories, Randall Schweller on neoclassical realism, and Jack Snyder on the empirical aspects of normative research in areas like ethnic conflict. Part II contains a mixture of commentary on the aforementioned applications, critiques of MSRP, and arguments for alternative rationalist models of scientific development provided by David Dessler, Roslyn Simowitz, John Vasquez, and Andrew Bennett. The coherence of the book derives in part from strong editorial guidance in the overview of MSRP, but also a conference held in Arizona in 1999, where the ideas that shaped this book were debated in spirited fashion.

This book is a fine example of methodological work that bridges the quantitative-qualitative divide. The model derived from Lakatos’s dense prose is clearly operationalized by Elman and Elman in Chapter 2 to guide subsequent applications (see also the handy brief guide on pages 19–20). The following chapters describe key elements of theories or successions of theories contained in well-known IR research programs. The contributors charged with applying MSRP reconstructed their research programs adductively by comparing observations with the model.

In keeping with the adductive method, many of the contributors questioned the operational definitions created by the editors as they were piecing together evidence and theory in their respective research programs. There were disagreements over the editors’ preferences for heuristic novelty as the operational definition of novel facts, with Moravcsik and Bennett making strong cases for the use of background theory novelty. Most contributors noted the difficulty of identifying all of the elements of a scientific research program in a nonarbitrary manner. Determining what counts as the hard core of a research program is complicated by Lakatos’s

suggestion that it may evolve over time. This also complicates identifying the negative and positive heuristics and consequently judging whether an interprogram or intraprogram problem shift has occurred (DiCicco and Levy, 151). Thus, even with the most precise operational definitions of the elements of MSRP to date in political science, the various contributors often struggled to match them with their observations. As a result, contributors variously found that their observations fit the model of MSRP (Keohane and Martin, DiCicco and Levy, Ray), that their observations did not fit the model (Schweller), and that while their observations may have fit the model, this still did not satisfy them that MSRP was the best way to assess their research program (Moravcsik, Walker).

Some contributors identified aspects of their own research program as degenerative, but argued that on the whole the research program itself was still progressive. Simowitz (407) was critical of these conclusions, arguing that Lakatos provides us with no way to aggregate progressive and degenerative aspects of a research program into an overall evaluation of the program. Several contributors argued that other research programs (i.e., realism) were degenerative or even falsified by Lakatos's criteria in comparison to their own (Moravcsik, Ray). Yet, as DiCicco and Levy (155) note, the research programs analyzed in this book are characterized by a kind of selection bias—all of the research programs reviewed have been successful at enduring through time. While we may be able to reconstruct these programs using MSRP, we can't answer if progressive research programs are more likely to endure than degenerative programs in the absence of negative cases. We also don't know if progressive programs may fail for other reasons not considered by MSRP.

Several issues arose in the book that deserve further mention. First, what is the most appropriate "unit of appraisal" for MSRP? There was little agreement on this issue among the contributors. Moravcsik argued that MSRP may be most useful for evaluating very large units like rationalism and constructivism. Ray similarly suggests that the units of appraisal may be large, but he characterizes the units in terms of systemic versus behavioral research programs. Walker suggests that Laudan's problem-solving approach is more consistent than Lakatos's with respect to focused research programs that rely on middle-range theories. Bennett concurs and advocates more attention to problem-driven research programs rather than the "isms."

A second issue not directly addressed by the contributors is whether research programs that em-

ploy particular methods may be privileged by MSRP's criteria. DiCicco and Levy remind us that Lakatos defines the hard core in terms of theoretical assumptions, while completely ignoring methodological assumptions. As many of the contributors note, aspects of empirical progress such as better operational indicators are not considered by MSRP. This kind of progress is often a hallmark of quantitative research programs. Yet, as Dessler points out, MSRP also fails to consider improved explanations of history as indicative of progress, which puts many qualitative scholars at a disadvantage. Multimethod research is not privileged either as DiCicco and Levy and Simowitz emphasize that corroborations of a fact through repeated tests with different operational measures, or different empirical domains, time periods, or different research designs are not considered evidence of progress by MSRP. In sum, no particular methodological approach seems to have any advantage over others in the use of MSRP.

Finally, Lakatos's MSRP is as close to a common metric that we have for comparing progress in both qualitatively and quantitatively oriented research programs. Despite DiCicco and Levy's (156) concern that Lakatosian theory was being subjected to a version of Popper's naive falsificationism (since MSRP was pitted solely against the data rather than against rival meta-theoretical frameworks) the consensus of the contributors was that Lakatos was a useful place to start in judging scientific progress in IR. A number of contributors called for explicit future comparison with other rationalist models. Others began to incorporate such comparisons in their analysis in this book. Moravcsik and Walker both suggested Laudan may more appropriately capture how their research programs have progressed. Bennett also identified three additional post-Lakatosian rational approaches to scientific progress that may hold promise: the Bayesian approach, the error-statistical school of thought, and naturalist approach associated with Laudan and others. Much like MSRP, these post-Lakatosian approaches may also be applied to research programs that use qualitative and/or quantitative methods.

Overall, this edited volume has much to offer the discipline. It is a clear and concise introduction to Lakatosian meta-theory and its application in IR. The contributors' efforts to identify the key elements of MSRP in their IR research programs may serve as models for similar comparisons in other subfields in Political Science. The applications demonstrated that rationalist meta-theory can equally inform research programs characterized by qualitative and/or quantitative methods. As a discipline we often take stock of

the sum of our work, but usually without reference to explicit meta-theoretical principles. This book provides us with an articulated rationalist approach to such endeavors and prompts us to think about other approaches that may similarly allow us to assess progress in our fields of study.

Cameron G. Thies, *University of Missouri-Columbia*

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

doi:10.1017/S0022381607080267

Concepts are central to social science, yet there are relatively few books devoted to helping social scientists use concepts appropriately. Gary Goertz has provided significant assistance in this regard with *Social Science Concepts: A User's Guide*. This is an advanced book that hits the ground running. Chapter 1 introduces the complex and provocative arguments in the book; it does not provide an introduction to the literature. In this regard, the term “user’s guide” in the subtitle is a bit misleading. *Social Science Concepts* is not a full-service guide on the formation and use of concepts. It is a cutting-edge book that makes specific arguments about the relationship between concepts, measures, and cases. It provides a common framework for both quantitative and qualitative researchers, which makes it a useful foundation for mixed-methods research. While the logic applies to both qualitative and quantitative scholars, Goertz argues that each is prone to different types of errors. For quantitative researchers, the basic message is to pay closer attention to the relationship between concepts and measures. For qualitative researchers, the basic message is to pay closer attention to the relationship between concepts and cases.

Goertz is concerned with the ontological properties of concepts, not the linguistic properties. Hence, this is not a “user’s guide” for those studying the ways concepts are used in ordinary language. Readers seeking a broader overview of, and introduction to, the literature on concept formation would be better served by John Gerring’s *Social Science Methodology: A Criterial Framework* (2001). These two books, however, would make good companions in a graduate course on research design, because Gerring’s framework allows us to place Goertz’s argument in context. Gerring argues that we should think carefully about the criterial trade-offs involved in concept formation, because these criteria have major impli-

cations for research design. Goertz explicitly favors two of Gerring’s eight criteria for conceptual goodness: analytic utility (the usefulness of a concept within a particular theoretical context or research design) and field utility (the usefulness of a concept within a field of closely related terms). In Goertz’s words, “The central attributes that a definition refers to are those that prove relevant for hypotheses, explanations, and causal mechanisms” (4). Goertz also implicitly favors validity (because he takes an ontological view of concepts) and operationalization (because of his empirical orientation). He does not attempt to appeal to readers who might favor other criteria for conceptual goodness, particularly linguistic criteria, such as resonance (familiarity) and contextual range (usefulness across linguistic contexts). This is a “user’s guide” for positivists.

The first part of the book focuses on the relationship between concepts and measures. As Goertz notes, methodologists have typically focused either on concepts or on indicators; relatively few have thought systematically about the relationship between the two. Thus, Goertz seeks to violate his own Second Law: “The amount of attention devoted to a concept is inversely related to the attention devoted to the quantitative measure” (2). His discussion of measurement, however, is not strictly quantitative; it also covers what some would call qualitative coding schemes, because he considers dichotomous variables to be special cases of continuous variables.

Goertz begins by acknowledging the important groundwork on concept formation laid by qualitative methodologists Giovanni Sartori (on classical categories and conceptual stretching) and David Collier (on family resemblance categories). He then quickly builds upon their foundation with his own quantitatively oriented framework, which he labels causal, ontological, and realist. “It is an ontological view because it focuses on what constitutes a phenomenon. It is causal because it identifies ontological attributes that play a key role in causal hypotheses, explanations, and mechanisms. It is realist because it involves an empirical analysis of the phenomenon” (5). In these regards, he carefully distinguishes his approach from the factor analytic approach, in which there is a causal relationship between concepts and indicators (14–16). Democracy, he argues, is defined (in part) by the presence of elections; hence, elections constitute democracy, they are not caused by democracy.

In Chapter 2, he formalizes the mathematics of concept structures. Goertz argues that most concepts

are either classical or family resemblance. Since classical concepts are based on necessary and sufficient conditions, they can be formalized by using the logical operator “AND” (e.g., democracies must have characteristics A, B, and C). Family resemblance concepts, by contrast, are based on sufficiency conditions; hence, they can be formalized by using the logical operator “OR” (e.g., democracies must have either characteristic A or B or C). Stated differently, family resemblance concepts have substitutable characteristics; classical concepts do not. While the logic is straightforward, Goertz demonstrates in Chapter 4 that numerous scholars have failed to follow it when measuring democracy. Most scholars have defined democracy in terms of necessary conditions (the logical “AND”) but have measured democracy using substitutable characteristics (the logical “OR”). In Goertz’s terms, this type of research produces low “concept-measure consistency” because the measures do not reflect well the basic structure of the concept (95). Some might recognize this as a threat to validity, though Goertz does not use that language.

Significant errors then occur when researchers combine these measures via addition or averaging. Researchers assume these measures can be combined because they are highly correlated, and thus treat the measures as substitutable, even though they have defined the concept in nonsubstitutable terms. In other words, the mathematical structure of the combined measure is not consistent with the theoretical structure of the concept. Goertz provides an additional extended example of “dyadic democracy” in Chapter 5 (coauthored with William Dixon). The mathematical formalizations and technical advice in Part One (Chapters 2–5), including discussions of fuzzy logic and set theory, are much more nuanced than can be covered here. Readers will have to discover these insights themselves.

The second part of the book (Chapters 6–8) addresses the relationship between concepts and case selection. These chapters (all coauthored) focus primarily on differentiating positive, negative, and irrelevant cases. Positive cases are those the researcher seeks to explain (e.g., revolutions); negative cases are controls used to test the theory in question (e.g., revolutions that could have occurred, but did not); irrelevant cases are not useful for testing the theory in question (e.g., revolutions that did not occur AND were highly unlikely to have occurred).

Chapter 6 (coauthored with J. Joseph Hewitt) provides guidance on how to separate positive and negative cases by using classical categories. This is relatively straightforward, in that requiring addi-

tional necessary conditions restricts the cases that qualify as positive or negative cases (and thereby reduces the gray area in between). This chapter also demonstrates how selection biases can be introduced by following this guidance. Unfortunately, the chapter does not advise how to separate positive from negative cases when dealing with family resemblance categories.

Chapter 7 (coauthored with James Mahoney) provides guidance on how to weed out irrelevant cases by using the Possibility Principle. This chapter is a revised version of a previously published article in the *American Political Science Review* (98: 653–69). “The Possibility Principle states that the negative cases should be those where the outcome has a real possibility of occurring—not just those where the outcome has a nonzero probability” (179). They argue that statistical researchers commonly consider all cases as relevant for testing, based on “the belief that excluding cases as irrelevant entails the loss of potentially helpful information” (182). Yet bigger sample sizes are not always better, since adding irrelevant cases artificially inflates the number of observations that confirm a theory. “In effect, this practice can make a false or weak theory appear much stronger than it really is” (183). Chapter 8 (coauthored with Hewitt) extends this argument to apply the Possibility Principle to the selection of populations, not just negative cases.

The third part of the book contains a single chapter (coauthored with Mahoney) on the use of concepts in theories. It introduces the idea of two-level theories and ties it back to earlier parts of the book on three-level concepts (the basic, secondary, and indicator levels). Part Three is necessarily the least developed part of the book, because it contains only one chapter. As the concluding chapter of the book, it serves to open paths for future work on methodology, rather than to summarize the book. Readers seeking to tie all the pieces of the book together should do so by rereading Chapter 1.

In sum, *Social Science Concepts* is an important addition to the literature on concept formation, measurement, and case selection. It is not a complete user’s guide because it appeals to a particular subset of social scientists. It is, nonetheless, a user’s guide because it gives attention to what users should do if they accept the basic assumptions of the book and provides numerous do’s and don’ts along the way. Many of these do’s and don’ts are summarized as bullet points in Chapter 2. It would have been useful to continue these bullet-point summaries throughout the book, but this is to quibble with the organization

of an otherwise engaging, thought-provoking, and much needed user's guide on social science concepts.

Craig W. Thomas, *University of Washington*

Interpretation and Method: Empirical Research Methods and the Interpretive Turn. Edited by Dvora Yanow and Peregrine Schwartz-Shea. (M.E. Sharpe, 2006)
doi:10.1017/S0022381607080292

Unlike the other volumes on qualitative methods being reviewed in this issue of *JOP*, Yanow and Schwartz-Shea manifest no interest in making qualitative methods useful or digestible for the mainstream. Instead, they argue, in their introduction and conclusion, in their introductory pages to each section, and in their own individual chapters, that interpretivist methods are incompatible with mainstream social science, and the latter more often "subjugates" qualitative methods, than collaborates with them.

Meanwhile, there is a second book going on within the volume, the book that comprises the other 17 substantive chapters. Many of these contributors, unlike their editors, are apparently interested in a conversation with the mainstream, while preserving the integrity of interpretivism in the process. They offer many useful ways in which interpretivist methods, once made explicit in their actual employment, can be understood as consistent with more mainstream methodological approaches without becoming subordinated to, and hence, distorted by, the positivistic priors of mainstream social science.

The chapters and introductions by the editors are sufficiently provocative and challenging to deserve a read by any curious mainstream scholar, but the chapters by the contributors are likely to enjoy and profit a still broader audience because of their efforts to make interpretivist methods more accessible and comprehensible to any and all researchers who would like to investigate social phenomena from the perspective of understanding the meanings of actions, rather than assuming that these meanings are self-evidently obvious.

To begin, Yanow and Schwartz-Shea object to the very term "qualitative" methods, pointing out that it has too often been reduced to merely small-n research or comparative case-studies, the latter most frequently not adopting a constructivist ontology or an interpretivist epistemology. This leads to one of the editors' more controversial claims, one not shared by some of the authors in the book, viz., that what should drive one's methodological choice is one's

epistemological and ontological priors, not one's topic of research. Surely it is true that if one has such priors, methodological choices necessarily follow. But it is an open question whether in fact everyone should have such priors, or whether one should allow one's methods to derive solely from the nature of the question being asked. I think their larger point, however, is valid, viz., that the mainstream's positivistic assumptions are so deeply internalized that issues of epistemology and ontology don't even arise, thereby privileging by silence and inattention, the positivist default. If everyone at first just considered how they thought knowledge was validated, and what they thought reality was, perhaps nonstandard approaches, such as interpretivism, would often follow.

Joe Soss was surprised to find out that his work was called interpretivism and that he was expected to be an interpretivist "whose worldview is defined by a particular epistemological and ontological paradigm." He noted that "packs of scholars take sides in longstanding philosophical disputes..." Reflecting on his own work, he found "it hard to square such accounts with experience. Most of my work is question driven," he concluded (130; see also Shehata and Bevir).

A second provocative claim by the editors is that interpretivism should have its own criteria for assessing the validity of its claims and not import them from mainstream positivism. This strikes one as a creditable position, although many authors in the volume appear not to adopt it, for they advance quite a long list of methodological mechanisms they call interpretivist that would not at all be misplaced in a mainstream methods primer. These include an explicit sampling strategy, defensible case-selection criteria, elaborating "falsifiable" propositions, concern with representativeness in one's empirical evidence, researcher effects, investigation of possible spuriousness of results, validating transcriptions with interview subjects, variation of values of hypothesized variables, establishment of scope conditions for theoretical claims, consideration of alternative explanations, among many others along the way (87, 98–09, 104–109, 116, 133, 159, 260, 294–99, 385, and 390).

So, while I think the broader point is crucial, viz., that interpretivist approaches do in fact preclude the use of many mainstream methods, that doesn't mean that all must be rejected. Moreover, adoption of many of them have the effect only of making interpretivist truth claims that much more convincing, both to the mainstream, and interpretivists themselves. For example, I have no idea why interpretivists should reject

validity and reliability as criteria for evaluation of their work. Why shouldn't interpretivists want their conceptualization and categorization of the meanings of their subjects' actions and practices to correspond to how these subjects themselves understand them? Isn't this the bedrock of interpretivism: allowing the intersubjective world to speak? Indeed, several of the authors argue the need for establishing the validity of their interpretations by giving transcripts back to subjects for their perusal and approval.

As for reliability, why wouldn't an interpretivist researcher be happy to see another researcher come along, look at the same sample of texts, or observe the same shop floor, and come away with similar conclusions about the nature of the discourse and practices she observes? What kind of violence does this do to interpretivist priors?

One principle of interpretivism highlighted by the editors is reflexivism, understood as the inevitable involvement of the researcher in the production of her data, what the mainstream would readily understand as bias. But what the mainstream considers to be bias that can be controlled for, the interpretivist believes is just how it is, and may merely be taken into account when reporting one's results, or indeed embraced, as a necessary part of social phenomena, the unavoidable conflation of facts and values. But here I think the editors somewhat idealize or simplify the interpretivist world, as not all practitioners explicitly situate themselves in their research. This is far more characteristic of ethnography, participant-observation, and Gadamerian hermeneutics, than it is for textual analysis, for example. Indeed, only a minority of scholars reporting how they do their work in this volume highlight how they situate themselves in their research, and what effects it has on the truth claims they make. In other words, it is just one stream of interpretivism, not a defining element.

Reflexivism also implies a critical approach to one's work that is largely absent from the mainstream, viz., the unquestioning acceptance of concepts, theories, variables, and histories, as if they have no social origins or are not instrumentalities in political plays of power (Oren, 215–27) Two examples from International Relations spring to mind. The first is John Lewis Gaddis's "long peace" which is used to describe the Cold War. Its analogous term is Kenneth Waltz's "bipolar stability" during the Cold War. What makes these terms odd, at second glance, is that the ideas of peace and stability are applied to an era which saw 20 million dead from Afghanistan to Vietnam, Guatemala to Kampuchea. Surely, schol-

ars should interrogate a theory that calls widespread carnage peace or stability? Or should we just test those variables in other contexts?

One very important contribution of both the editors, and the book itself, is their repeated injunction that interpretivists should make their methodological choices and processes more explicit in their work, so as to increase confidence in its claims for validity, its prospects for replicability in, and perhaps even generalization across, other cases. Moreover, by making their methods explicit, interpretivists are likely to convince many more mainstream scholars about the systematic and rigorous attention interpretivist researchers give to many mainstream methodological issues, such as case selection.

Finally, the value of the many substantive chapters should not be underestimated. I think many mainstream scholars crave examples of methodologically self-conscious interpretivist research, having become tired of all the meta-theoretical mud slinging, but not having lost the desire to see how the work is actually done, and then perhaps realizing how they might themselves adopt some interpretivist practices in their own work, having been already convinced that interpretivism's commitment to an intersubjective reality is probably a most reasonable ontological claim, but having been put off by years of distracting attacks on their own methodological choices as theoretically ignorant and politically reactionary.

Robert Adcock's chapter, for example, elaborates an interpretivist critique of Mahoney and Rueschmeyer's work on comparative historical analysis that points out the costs of a mainstream conceptualization of that approach. Dean McHenry Jr.'s chapter explores how an interpretivist approach could correct a frequently used data base of political demonstrations in the world, a data base that omits, in just the case of India, nine kinds of political activism that don't fit into the a priori categories assigned by the mainstream analysis. Perhaps what should be most encouraging, to the mainstream and interpretivists alike, is Maynard-Moody and Musheno's receipt of a National Science Foundation grant to collect stories from state employees about their daily lives at work.

In sum, both books within this volume are well worth reading, although the one written by the individual contributors is more likely to help you win that NSF funding.

Ted Hopf, *Ohio State University*

Fuzzy-Set Social Science. By Charles C. Ragin. (University of Chicago Press, 2000.)
doi:10.1017/S0022381607080309

Charles Ragin would have been hard pressed to come up with a book that is more likely, at first glance, to deflect the interest of large-N political methodologists and practitioners. It has “fuzzy” in the title, and to researchers interested in precision, that can’t be good. It takes seriously the proposition that necessary and sufficient conditions exist. Its foil, throughout the first chapter especially, is “conventional quantitative research,” whose practitioners are blind to so many aspects of the social world that even a generous reader would have to conclude that they are rather dim.

Setting the book aside for these reasons would be a very substantial mistake, however. Ragin’s focus has much more to do with analyzing causal complexity than with fuzzy sets per se; indeed, the latter are utilized in pursuit of the former. And any quantitative researcher who has contemplated using interaction terms in a simple regression equation has already admitted that causal complexity—in this case, the contingency of the effect of one variable on the level of another—may happen. The main question is not, therefore, whether it exists, but rather how we study it.

Ragin’s answer to that question is an extension of the brilliantly innovative Qualitative Comparative Analysis (QCA) technique that he laid out in his 1987 book, *The Comparative Method*. QCA’s emphasis on Boolean algebra as a remarkably efficient means of understanding causal complexity, coupled with its utility in even very small-N situations, made it immensely attractive to qualitative researchers. Quantitative researchers were less enamored of its deterministic formulations and substantial sensitivity to mismeasurement.

Fuzzy-Set Social Science reflects its author’s considerable reflection on these issues in the intervening 13 years. On offer is a new and improved QCA—fuzzy-set QCA, or fsQCA—that replaces the strict either-or formulations of QCA with the more nuanced concept of degree of membership in a set. Here, again, differences with quantitative research are easy to exaggerate: anyone who has formulated a survey question asking whether respondents are “very religious,” “somewhat religious,” “not very religious,” or “not religious” is already thinking in fuzzy-set-theoretic terms. The main difference, it seems, lies in how the answer is then quantified: a mindless 4-3-2-1 coding, Ragin argues, does not reflect the researcher’s substantive knowledge of the meaning of the answer. A better measure of religiosity might be 1-0.8-0.2-0, or

something similar, depending on the exact wording of the question. By the same token, a measure of atheism would not necessarily simply be the inverse of the measure of religiosity: even a little religiosity would disqualify the respondent from the set of atheists, so a better coding might be 0-0-0.2-1. One can argue the exact numbers, or how they should be derived, but the larger point is that we’re actually speaking the same language.

The first step in getting from coding membership in fuzzy sets to understanding causal complexity relies on the very important insight that absences of data can be as important as, if not more important than, their presence. (The absence of wars among democratic states, for example, has spawned a massive scholarly literature and has even been noticed in the policy realm.) These absences tell us quite a bit, according to Ragin: a dearth of observations in the upper-left half of an X-Y scatterplot suggest a relationship of causal necessity, whereas a dearth of observations in the lower-right half suggest a relationship of causal sufficiency. Again, the language makes the difference seem larger than it is, as the “dyadic democracy is sufficient for peace” formulation illustrates nicely.

The remainder of the method for understanding causal complexity comes down to addressing the problem of combining causal conditions. For example, one might argue that so many young people enter law school each year that the market is nearly saturated, and only young attorneys who are naturally gifted and who work very hard will have very successful careers. The argument implies that there should be a triangular relationship between membership in the set of gifted and studious people and membership in the set of successful lawyers, with the former establishing the limit to the latter: few if any lawyers should be more successful than they are gifted and studious. But even if we can measure degree of giftedness (using information from college GPA, rank of college, SATs, LSATs, etc.) and amount of time studying and convert those meaningfully into degrees of membership in the sets of gifted people and studious people, we are still left with the problem of aggregation: If Jones’ membership in the set of gifted people is 0.8 and her membership in the set of studious people is 0.5, what is her membership in the set of gifted and studious people? Ragin’s response is that the minimum defines the conjuncture of the two sets: Jones’ membership in the set of gifted and studious people is 0.5. (Similarly, Jones’ membership in the set of gifted OR studious people would be the maximum of the two, or 0.8.)

At this point, even sympathetic quantitative researchers may well rebel, because both the cutoff value for necessity and sufficiency ($X = Y$) and the rules for combining multiple causes ($\min(x_1, x_2)$ and $\max(x_1, x_2)$) are pegged by assumption rather than allowed to vary as the data demand. Ragin recognizes this point in the first instance, introducing a “fuzzy adjustment” (how he could resist “fuzzy fudge factor,” I don’t know) to move the diagonal up or down a bit to capture cases that are near the line. But the use of minima and maxima is a dauntingly strong assumption. It rules out the possibility that one of the two factors plays a bigger role in producing the outcome. Though Jones is more gifted than studious, for example, studiousness may contribute more to success than giftedness in general: perhaps $\min(\text{giftedness}, 2 * \text{studiousness})$ would better capture the quantity relevant to success. Not allowing any weighting of the variables will constitute a very substantial handicap to people who believe in estimating coefficients based on the data.

Opinion will probably be somewhat more divided on the sole statistical component of the procedure, a binomial test to determine whether or not the proportion of “successes” (cases below the $X = Y$ line for necessity, or above for sufficiency) is significantly greater than a given cutoff—0.65 for “usually necessary” and 0.80 for “almost always necessary” for example. Aside from the fact that, even to those who grant a probabilistic conception of necessity, one exception in five may well stretch the concept of necessity beyond acceptable limits, there is a purely statistical point that should be mentioned. The 80% benchmark is chosen because it represents a surprisingly high percentage relative to what we would expect by chance. But the maximization and minimization operations on the data have an impact on what we

would expect by chance: taking the maximum of three variables whose observations are drawn purely at random from a uniform distribution on the unit interval will produce a new variable that is heavily skewed toward the high end of the scale. If we plot this new variable against another variable, Y , with observations again drawn at random from a uniform distribution on the unit interval, it would not at all be surprising to find that 80% of them fall below the $X = Y$ line. (When I tried it myself, with $n = 20$, the average was 75% across 100 trials.) One could surely tailor more reasonable benchmarks based on expected outcomes for random data given the number of observations and the operations that produced X , and the procedure would benefit from it.

Finally, given Ragin’s focus on sets it seems surprising that he has nothing to say about what researchers in the set of *unconventional* quantitative scholars have to say about complexity. Given that conventional quantitative scholarship seems to consist of linear regression and correlations, this is a substantial omission. Only interaction terms are mentioned as plausible quantitative analogs to fsQCA, and their shortcomings—the need for a large N , the dangers of multicollinearity—are exaggerated.

In all, the book constitutes a very innovative approach to an interesting problem, and the elements of that approach are more familiar than a quantitative audience might suspect. That is not to say that it will command universal agreement, but it should cause even the most conventional quantitative scholar to question the conventional wisdom in many ways. Qualitative scholars, in turn, could benefit from their reactions. In short, it is a book that very much deserves to be talked about, and that is high praise indeed.

Bear Braumoeller, *The Ohio State University*