

Qualitative & Multi-Method Research

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

Contents

Symposium: John Gerring's *Social Science Methodology: A Unified Framework* (2012)

<i>Commentary on John Gerring, Social Science Methodology</i> Allan Dafoe	2
<i>Review and Critical Comments on Gerring's Social Science Methodology</i> Derek Beach	4
<i>Does Set-Relational Causation Fit into a Potential Outcomes Framework? An Assessment of Gerring's Proposal</i> Carsten Q. Schneider and Ingo Rohlfing	8
<i>Concepts and Description in Gerring's Social Science Methodology, 2nd ed.</i> Ariel I. Ahram	14
<i>Against Methodological Unity: A Critique of Gerring</i> Patrick Thaddeus Jackson	19
<i>Response: Fear and Loathing on the Methodological Trail</i> John Gerring	21

Symposium: Research Openness in the Digital Age

<i>Qualitative Data Access and Research Transparency</i> Colin Elman	28
<i>Closing the Infrastructure Gap: Qualitative Data Archiving</i> Diana Kapiszewski	31
<i>Active Citation and Qualitative Political Science</i> Andrew Moravcsik	33
<i>Undergraduate Scope and Methods Courses: Is There Room for Qualitative Methods?</i> Cameron G. Thies	37

Announcements

<i>APSA Short Courses and Section Panels</i>	39
--	----

APSA-QMMR Section Officers

President: Gary Goertz, University of Arizona
President-Elect: Lisa Wedeen, University of Chicago
Vice President: Jeff Checkel, Simon Fraser University
Secretary-Treasurer: Renee de Nevers, Syracuse University
QMMR Editor: Robert K. Adcock, George Washington Univ.
Division Chair: David Waldner, University of Virginia
Executive Committee: Elisabeth Wood, Yale University
Stephen Hanson, The College of William and Mary
Giovanni Capoccia, University of Oxford
Benjamin Smith, University of Florida

Letter from the Editor

Robert Adcock

George Washington University
adcockr@gwu.edu

With this issue our section's newsletter moves into its tenth year. The creation of the section a decade ago provided an expanded base within the APSA to address the premises, practices, and prestige of qualitative and multi-method research. This development dovetailed with a broader concern among political scientists to better appreciate the full spectrum of varied methods used in the discipline. Although we are at least ten years into this conversation, there is no sign of any lessening of interest. Indeed, if anything, attention is picking up. Take, for example, the books that have appeared in the few months since the newsletter's last issue. First, the Cambridge University Press series "Strategies for Social Inquiry" was inaugurated with a revised and extended second edition of John Gerring's *Social Science Methodology*. In the months ahead in 2012 we can expect to also see in this series Michael Coppedge's *Democratization and Research Methods*, Thad Dunning's *Natural Experiments in the Social Sciences*, and Carsten Q. Schneider and Claudius Wagemann's *Set-Theoretic Methods for the Social Sciences*. Second, the Routledge Series on Interpretive Methods was just inaugurated with Peregrine Schwartz-Shea and Dvora Yanow's *Interpretive Research Design*. Third, alongside these two exciting series, a wider trend of publications in method and methodology from multiple major presses is evident in Clarke and Primo's *A Model Discipline* (Oxford University Press, 2012), and will continue in the months ahead with Gary Goertz and James Mahoney's *A Tale of Two Cultures* (Princeton University Press, 2012), as well as Derek Beach and Rasmus Brun Pederson's *Process Tracing Methods* (University of Michigan Press, 2012).

One mission of this newsletter is to help keep section members abreast of a quickly growing and changing literature. The current issue starts with a lively symposium of diverse commentaries on Gerring's new edition, which inaugurated the Cambridge series, and a symposium is planned for the next issue on Schwartz-Shea and Yanow's book inaugurating the Routledge series. While it is not possible to engage every new work in the limited space of the newsletter, as editor I hope to cover as extensive and diverse a selection of new works as possible. I pulled together the list of new works in my opening

Letter from the Editor continued on page 43

**Symposium: John Gerring's *Social Science Methodology*:
A Unified Framework, 2nd ed. (Cambridge University Press, 2012)**

*Commentary on John Gerring's
Social Science Methodology*

Allan Dafoe

Uppsala University, Sweden
Allan.Dafoe@pcr.uu.se

Social scientists tend to cluster in groups that share methodological commitments.¹ Formal theorists emphasize parsimony, precision, and logical coherence; more policy-relevant theorists tend to privilege richness, accuracy, and usefulness. Advocates of large-N statistical analysis place greater weight on generality, objectivity, and formal estimates of uncertainty; area studies scholars emphasize context-bounded generalizations and historical richness. Experimentalists prioritize clean (i.e., exogenous) treatment assignment and bounded but credible estimates of average causal effects; others prefer to study large thorny questions using designs that admittedly suffer from problematic biases. As more scholarly conversations take place within, rather than across, these methodological groupings, there is a risk that each group will come to hold its respective weighting of methodological criteria as superior and will forget the methodological trade-offs that led other domains of social science to prioritize other criteria.

Gerring's *Social Science Methodology* (SSM) offers an extremely valuable vocabulary and set of reflections to remind scholars of the broader objectives of social science, of the many different criteria that characterize good social science, and of the various trade-offs inherent in different kinds of social scientific research. Gerring's insight is to realize that even the most extreme methodological camps will generally agree in the abstract about characteristics of good social science. Social science "aims to be cumulative, evidence-based (empirical), falsifiable, generalizing, non-subjective, replicable, rigorous, skeptical, systematic, transparent, and grounded in rational argument" (Gerring 2012: 11). *Ceteris paribus*, an improvement in one of Gerring's many criteria will lead to better social science. Methodological disputes amongst most social scientists do not arise over fundamental disagreements about what criteria are important, but rather different solutions to the trade-offs between these criteria. Scholars disagree about the optimal weighting over different criteria, not the criteria themselves.

Any scholar who tries to remind various methodological camps of the benefits of approaches pursued by others takes a risk. Such a message induces dissonance and signals a lack of adherence to the preferred criteria of the respective camp. However, such a message is crucial to facilitate productive conversation across methodological groups, and to improve the use and evolution of methods in the field at large. Gerring does better than most at appraising the needs of vastly differ-

ent corners of social science. He articulates his methodological framework in a way that takes their respective priorities and challenges seriously, without sacrificing the goal of building a unified set of criteria that explicitly elevates certain modes of analysis when they are possible. For example, Gerring reminds non-experimentalists of the power of random treatment assignment on large independent samples to give credible causal inferences, while also reminding experimentalists that their methods remain unable to speak to many important questions. According to Gerring, every research program should be expected to employ the *best-possible* methods for their problem; methodologists should not push their preferred criteria on another research area without sufficient appreciation of the associated trade-offs in that domain.

Simplified, Functional Typology of Criteria

It is a tremendous achievement to articulate a practical, largely exhaustive list of criteria for social science. One shortcoming, as is almost inevitable for a task such as this, is that the list of criteria is sprawling and imperfect. Inevitably I found myself collapsing, splitting, and rephrasing criteria to have a framework that made more sense to me. For example, I found it helpful to group and modify Gerring's criteria for good theory (see his Table 3.1) under three higher-level criteria:

- (1) Testability²: precise (bounded) and strong, high explanatory power, elaborate and risky predictions, general (and causal), and minimal researcher degrees of freedom (parsimony, precision, explicitness, logical coherence).
- (2) Consistency³: external consistency, commensurability, logical coherence.
- (3) Usefulness: general, causal, relevant, important, novel.

These high-level criteria of testability, consistency, and usefulness follow from a simple understanding of the purpose of criteria: to increase the probability of achieving useful knowledge. Theories that are more testable—by which I mean subject to empirical evaluation—will be subject to more stringent selection through scientific research; theories that are more consistent with other (corroborated) theories are more likely to be true; theories that are more useful will... well, be more useful. Perhaps in his 3rd edition Gerring could similarly work to build a simpler framework that is more explicitly derived from the core objectives and character of scientific research.

Just as the typology of criteria would be easier to process and employ if it had just a few top-level concepts that followed from first principles, *Social Science Methodology* as a text would be easier to digest if there were an opportunity to read a more concise summary of the argument. I understand that John Gerring may be in the process of writing such a book; if so it would provide a valuable addition. That being said, there is great value in the many digressions, footnotes, and reflections

in *SSM*; *SSM* is a useful reference text for methodology organized according to criteria and tasks, with many fascinating references and quotes. (I think students will especially appreciate the discussion of “Finding a Research Question” in Chapter 2, which is an important part of the research process that is neglected in most textbooks.)

Exhortation and the Advance of Methodological Standards

Science is a social process, and scientific standards advance through changing scientific norms. In addition to analyzing and wrestling with the trade-offs of different methods, a methodology textbook should exhort for better standards and point out strategies that are particularly effective. I would have appreciated it if Gerring allowed himself more opportunity to demonstrate examples of good social science, to articulate strategies for overcoming various trade-offs, and to insist on particular best practices. Methodologists should ponder the finer points of methodology, but, when possible, we should also provide clear recommendations and exhortations for practitioners.

Consider the example of norms of replicability. Gerring rightly points out that replicability and transparency are critical to social science. I would encourage him (and all methodologists), however, to more clearly denounce bad habits such as the failure to publish detailed replication files for quantitative research (where appropriate and possible). It is simply unacceptable that our top journals do not take the norms of replicability more seriously; legitimate exceptions such as protecting human subjects or providing authors with a time-window to exploit their collected data can be accommodated. Gary King jokes that a surprising number of replication datasets are stored in boxes that get lost when scholars move, and I can attest that in this day of cloud-storage, this source of data loss still plagues political scientists. Norms of proper behavior have the most force when they are clear and commonly accepted; it is part of our job as methodologists to clarify and advocate for norms of good research.

Most methodological recommendations are not so straightforward, but even still there are approaches and strategies that tend to work better than others: to score higher on most or all criteria, relative to other approaches for a given task. I encourage Gerring for his next editions to more directly share his perspective on which approaches are most effective. For example, there is an approach to social science referred to as *design-based inference*⁴ that I think represents a particularly promising future for social science. Design-based inference involves searching for and exploiting natural experiments: opportunities where some important factor was (conditionally) as if randomly assigned. Design-based inference involves statistical inference, and thus benefits from the objectivity and formal estimates of uncertainty of quantitative research. However, unlike much observational statistical research, design-based inference takes the assumptions required for causal inference much more seriously. Deep subject matter expertise is essential to find natural experiments and extensive qualitative research is invaluable to justify the key assumption that treatment assignment is (conditionally) as if random. Design-based

inference scores higher on many criteria, with few trade-offs, compared with many other approaches. However, one cannot simply tweak typical observational quantitative studies to score better on the various criteria associated with design-based inference; one has to have a fundamentally different disposition towards research. One has to be deeply familiar with the theory and examples of design-based inference, one has to be more open to letting one's specific research question be guided by design opportunities, and one has to have sufficient subject-matter expertise and qualitative research skills to find and evaluate potential natural experiments.

As an example, consider Saumitra Jha and Steven Wilkinson's (forthcoming) excellent study of the effect of combat experience for ethnic groups in India during WWII on later ethnic migration patterns arising from varying capability for political collective action. This design required insight about the haphazard nature of battalion deployment in WWII; in-depth qualitative evidence about how forces were enlisted, trained, and deployed; awareness of the potential value of certain data buried in historical archives; and expertise with the appropriate statistical methods. Out of this combination of methods and sensibilities comes the rarest of achievements in social science: a persuasive causal inference about an important phenomenon. Design-based inference marries multiple methods in a functional manner, making the whole much greater than the sum of its parts. There are trade-offs with different approaches in social science, but there are some approaches and strategies that are relatively superior for particular tasks on most criteria; perhaps in Gerring's 3rd edition or his simplified version he could exhort more strongly for those.

In this review I have held Gerring's *Social Science Methodology* to a high standard; if I hadn't I wouldn't have had much to say! *SSM* is a serious, important, and intelligent exploration of the common methodological ground across the breadth of social science. Scholars will want to turn to specialized textbooks for learning the particular methods and sensibilities of their trade. Gerring's *SSM*, however, should be of broad interest and general value; it demonstrates the methodological trade-offs arising in various corners of social science, it reminds us of our shared standards, and it makes a substantial contribution towards a unified framework for social science methodology.

Notes

¹ For helpful suggestions I thank Robert Adcock, Sophia Hatz, Mathilda Lindgren, and Nynke Salverda.

² By “testability” I mean *subject to empirical evaluation*, rather than a naïve falsificationism. I agree with Clarke and Primo (2007) that not every theoretical contribution needs to be judged by the extent to which it can be empirically evaluated; however, I think it is a mistake to think that empirical evaluation should not be a primary criterion of theory in general. Consider their example of Downs' spatial model of electoral politics: Even though its precise predictions may not be correct, the general predictions seem to have much empirical bite and the spatial model is especially valuable because it seems to correctly account for much political behavior.

³ The extent to which a theory is consistent with other well-supported theories.

⁴ Dunning (2010). Related approaches are referred to with the label *causal inference*, and in economics, the *credibility revolution*. See Freedman (2010), Sekhon (2009), Angrist and Pischke (2008).

References

- Angrist, Joshua D. and Jorn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Clarke, Kevin A. and David M. Primo. 2007. "Modernizing Political Science: A Model-Based Approach." *Perspectives on Politics* 5:4, 741–753.
- Dunning, Thad. 2010. "Design-Based Inference: Beyond the Pitfalls of Regression Analysis?" In *Rethinking Social Inquiry*. Henry E. Brady and David Collier, eds. Lanham, MD: Rowman & Littlefield Publishers.
- Freedman, David A. 2010. *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. New York: Cambridge University Press.
- Jha, Saumitra and Steven Wilkinson. Forthcoming. "Veterans, Organizational Skill and Ethnic Cleansing: Evidence from the Partition of South Asia." *American Political Science Review*.
- Sekhon, Jasjeet S. 2009. "Opiates for the Matches: Matching Methods for Causal Inference." *Annual Review of Political Science* 12, 487–508.

Review and Critical Comments on Gerring's Social Science Methodology

Derek Beach

Aarhus University, Denmark
derek@ps.au.dk

There are two parameters that can be assessed when critically reviewing the second edition of a well-established textbook. First, have the revisions made the book an even better tool to teach students about social science methodology? Second, have the revisions brought the book up to date with current debates, providing students with a state-of-the-art picture of the field? The following review concentrates on part III (causation) and part IV (conclusions) of the book. The review is written in light of the recent developments in the field of case study research by scholars such as Bennett, Checkel, Collier, Mahoney, and Waldner, among many others.

Is the Book a Good (Better) Teaching Tool?

As with all of Gerring's work, the book is written with exceptional clarity. There are numerous very lucid and accessible introductions to many difficult methodological questions. For example, the discussion of the difference between a background and a scope condition is excellent (pp. 205–206), as is the discussion of the importance of robustness tests when dealing with population-oriented observational designs (pp. 319–321). I was also pleased with Gerring mentioning that "thinking carefully about the theory and its relationship to the DGP [data-generating processes] serves to focus our attention in productive ways" (p. 322). The introduction to the distinction between causes-of-effects and effects-of-causes is

also excellent (pp. 333–334). Finally, clearly flagging that methodology and research design is all about trade-offs and choices is an important but all too neglected point in most textbooks.

Given that this is a critical review, I will however focus my attention on more problematic areas, concentrating on (1) whether the level of the book is appropriate for the target audience, and (2) whether the guidelines for research design are consistent.

The Book is Too Complicated on Many Topics

Unfortunately, on many topics the level of the book is too advanced for all but the most sophisticated student, and even here only those intimately familiar with what Mahoney (2008) terms population-oriented research methods will probably be able to understand (and find useful) all of the different terms. The best example of this excessive complexity is the terminology-heavy discussion of causality in chapter 9; all understood as variants of average treatment effects (ATE). What do students gain from knowing about variations of ATE such as SLATE, SITT, or LATE? While these terms might be important when discussing research designs within the experimental template, the utility of these forms of causal relationships for case-oriented research is questionable. What exactly is the value-added for qualitative scholars of going beyond King, Keohane, and Verba's introduction of the concept of mean causal effects as used in population-oriented research methods (King, Keohane, and Verba 1994), especially when many case-oriented researchers claim to be studying other types of causal relations than mean causal effects (Mahoney 2001, 2008)?

Another example of this excessive complexity is when different types of violations of causal comparability (confounders) are introduced (p. 251). Here a large range of confounders, including attrition, reputation effects, and testing effects, are introduced—all of which are very relevant for researchers adopting the experimental template, but for the majority of social scientists that use observational designs, many of these forms of treatment bias are less relevant. For example, if I am tracing a causal mechanism such as group think using observations in archives, bias in the form of noncompliance of respondents with experiment instructions is quite irrelevant.

The book also contributes to confusing students by using the term "necessary condition" in a much broader sense than just as a type of cause. For example, on p. 234 the book suggests that empirical covariation is a "necessary condition" of a causal relationship. Yet if the X in this causal relationship is theorized as a necessary condition, this terminology would lead a student to make the argument that covariation is a necessary condition to establish that a necessary condition exists (p. 234, 291). Here I would suggest a much narrower use of the term necessary condition, using it only to refer to causal conditions themselves, and not also as criteria for inferring causal relationships.

The Book Introduces a Level of Discretion that Can Confuse Students

Second, in order to avoid being overly dogmatic on the "unified approach to social science methodology," the author

attempts to keep case-oriented researchers on board by being pragmatic in several instances. However, the result introduces a level of discretion that students will be hard-pressed to make sense of. For example, the book on p. 365 posits that “increasing the N of a research design is desirable if all things are equal,” whereas the book suggests on the next page that sometimes “a few causal-process observations are more important than many dataset observations” (p. 366). How do we reconcile these two statements? When exactly is more better, and when is fewer better? Given the celebration of the experimental template and the use of a population-oriented understanding of causality (average treatment effect), most students will be left with the impression that big is (almost) always better (as in King, Keohane, and Verba 1994), and that we use case studies only when we are *unable* to do large-n research (preferably in the form of experiments).

Another example is when Gerring suggests on p. 323 that when conducting causal reasoning, “consult experts in the topic, that is, those with knowledge of the theoretical problem you are attempting to solve and the data-generating process you are attempting to interpret. We must recognize the importance of experience and wisdom in crafting and in vetting social science studies. Sometimes, the experts know best and their views should always be consulted.” While it is easy to agree that much population-oriented research on phenomena like the democratic peace thesis would have benefited from input from “experts,” if we are engaging in case-oriented research, we cannot just rely on “experts.” In most instances, to undertake in-depth case studies we will have to develop our own case-specific expertise; especially to understand what constitutes actual evidence of the existence of an underlying causal relationship or mechanism in a particular case. Should Gerring’s advice be read as suggesting that we should always rely on other case-oriented “experts” when evaluating evidence? What then if the “experts” have conflicting interpretations? And most critically, what if an “expert” disagrees with our interpretations of what caused specific events after we have done our own in-depth research on the case? While being pragmatic, this advice introduces a level of discretion that will be confusing for most students.

Does the Book Provide an Updated Picture on the State of Social Science Methodology?

Gerring celebrates experimental designs in the book. He writes that “The first question we tend to ask of a research design is whether a hypothesis can be tested in an experimental fashion, that is, with a randomized treatment” (p. 324). He goes on to state that “This suggests that one never really knows the truth of a nonexperimental result until one has conducted an experiment on the same question—a rather damning conclusion for those of us who wallow in the mire of observational data” (p. 325). Throughout the book, Gerring revealingly uses medical experiments as examples (pp. 384–385).

In reading the book, one gets the feeling that it is almost as if all of the heated debates about KKV’s *Designing Social Inquiry* have not taken place (King, Keohane, and Verba 1994; Brady and Collier 2010; Mahoney 2001; George and Bennett

2005; Freedman 2010). Indeed, Gerring in many respects takes an even more extreme position by lifting experimental designs up onto the pedestal of being *the* ideal-type of good (social) science, thereby ignoring the last ten years of developments within qualitative, case-study-related research—where the key point is that there is not a unified logic of social science (e.g. George and Bennett 2005; Brady and Collier 2010; Mahoney 2001, 2007, 2008; Ragin 2000, 2008; Waldner 2010).

The unified logic of social science that Gerring advocates is something that would be nice in the ideal world, but the heated debates on the nature of (social) science that have raged for millennia tell us that there is not one widely agreed upon standard of “good” social science and research design. Scholars disagree deeply about basic ontological questions such as the nature of causality or whether the complexity of social science phenomenon allows or precludes comparisons across large populations.¹ Brady for instance helpfully delineates four distinct ontologies of causation (manipulation, counterfactuals, regularity, and mechanisms). Other important distinctions are between deterministic and probabilistic understandings (Mahoney 2008). The key point here is that different ontological positions have different methodological implications when we keep in mind Peter Hall’s argument that our methodology needs to reflect the underlying ontological assumptions we make about the world (Hall 2003).

Yet while Gerring discusses competing “non-unified” ontologies of causation in the concluding part IV of the book, he assumes away the differences in the interest of creating a unified logic of social science. Within this unified logic, the experimental template is elevated above all other designs. However, it must be remembered that this method has at its ontological core a manipulation and probabilistic understanding of causality that is not shared by many social scientists. Elevating one distinct ontological combination has clear methodological implications for what is viewed as a proper research design in the book, and more importantly, what is not viewed as good social science.

Non-experimental methods are therefore seen as second- or even third-best designs, with an emphasis in the book on describing how we can compensate for their weaknesses in relation to true experiments. These weaknesses are seen as particularly acute for case study methods, with the book implicitly telling us that they are in effect third-best research designs. The first-best is the experimental design. When we cannot manipulate directly, Gerring advocates adopting second-best, large-n observational designs to assess average treatment effects across a population. When we cannot create a sufficient number of observations to engage in large-n analysis, we must settle for case studies as a third-best solution.

Two of the implications of the celebration of experiments are explored below. First, adopting a unified logic results in a narrowing of the types of research questions we can legitimately ask, and masks the comparative advantages of non-experimental methods. Second, elevating the underlying ontology of manipulation/probabilism produces misconceptions about other methods that hold different ontologies of causation.

Gerring suggests that all social science research designs aim to replicate the virtues of “true experiments,” and that the first question we ask is whether a hypothesis can be tested using an experiment (p. 324). Yet these suggestions simply do not make a lot of sense when seen from the vantage point of case-oriented research. If we are interested in studying why India is democratic despite everything we know from other cases about the relationship between economic development and democracy, is the first thing we ask ourselves “How can I test a hypothesis that could explain the deviance in an experimental fashion?”

Celebrating the experimental template and focusing our attention on assessing average treatment effects unfortunately reduces the scope of the research questions we can legitimately ask of the social world (pp. 376, 377). At times the book does open for a broader understanding (p. 226, where Gerring writes that “they also compromise a separate research agenda. We want to know why X causes Y, not simply the treatment effect of X on Y.” Yet these statements are inconsistent with the celebration of the experimental ideal that prevails in the rest of the book.

The choice of method should reflect the type of research question we are asking. An experiment might be a strong method to assess the magnitude of the average treatment effects of X upon Y. Research on framing of issues has, for example, benefited from experiments, where the information respondents receive is carefully designed and controlled to isolate framing effects from other factors (Gaines et al 2007). But if we want to investigate *how* X contributes to produce Y we should adopt a more appropriate method such as an in-depth case study analysis using process tracing (see below). If we are studying a complex phenomenon, there might be multiple causal pathways that could hypothetically produce Y. In this research situation QCA-related methods would have strong comparative advantages (Ragin 2000, 2008).

Instead of dogmatic prescriptions about “true experiments” as an ideal design, the guidance we need to offer our students is that (1) research questions determine the choice of method, not vice versa, and (2) they should choose the research method that offers the most bang for the buck in relation to a specific research question, i.e., it has the strongest comparative advantage in relation to offering analytical leverage in the particular research situation.

Misconceptions about Variants of Case Study Methods Due to the Unified Logic

Gerring’s book adopts the experimental template’s manipulation/probabilism ontology of causation as a unifying definition instead of acknowledging the importance that differences across ontologies of causation have to understanding the comparative advantages and disadvantages of different methods. While similar points to the following could be made about how Gerring describes other qualitative methods (e.g., congruence or QCA), I will concentrate on Gerring’s understanding of process tracing methods. In the following I define the method as

in-depth case studies that investigate causal mechanisms (see Bennett 2010; Beach and Pedersen 2012; Waldner 2010).

Gerring’s ontology of causation is one of manipulation, with counterfactuals playing an adjunct role. In contrast, process tracing arguably builds upon a mechanistic understanding of causality that is quite different from manipulation (see Beach and Pedersen 2012). At the core of the mechanistic understanding of causality are causal mechanisms, which are viewed by many as something much *more* than mere *intervening variables*. Instead, they are seen as a system of parts that together transmit causal forces from X to Y, contributing to producing an outcome (Bennett 2008a, 2008b; Glennan 1996, 2005; Bunge 2004; Hedström and Ylikoski 2010; Machamer 2004; Waldner 2010).

However, when seen from Gerring’s manipulation ontological position, causal mechanisms are just intervening or intermediate variables between X and Y (see also Gerring 2010). He describes in the book an example of what he terms a causal mechanism that links smoking and cancer, depicted as: (1) smoking \rightarrow tar, (2) tar \rightarrow cancer (p. 309). Yet it is evident from this example that he does not take mechanisms as seriously as the mechanistic ontology would suggest. What he describes here is an intervening variable (tar) between smoking and cancer, with no information about *how* tar actually produces cancer. When studying causal mechanisms we are not only interested in *whether* there is a causal relationship between X and Y, but more significantly in the theoretical process of *how* X contributes to produce Y. What then exactly is the causal mechanism whereby tar produces cancer? Unfortunately, Gerring black boxes the actual causal process by seeing a mechanism as an intervening variable. Yet to study it we need to unpack what is happening *within* the causal arrows themselves, theorizing a causal mechanism that can then be tested empirically.

Gerring’s cavalier treatment of mechanisms is also seen when he discusses them as confounders that should be controlled for (p. 297). But if we take causal mechanisms seriously, they are what occur *within* the actual arrows themselves. Gerring illustrates his points about controlling for mechanisms using an example of a mechanism whereby vouchers (X) might impact on Y (education attainment) through a quality of classroom instruction mechanism (M). He then suggests that we might want to control for quality of classroom when the mechanism is conditioned in a simple, one-stage covariational analysis (p. 297).

Yet a mechanism is not just an intervening variable whose effects can be isolated. It is something more; a system of individual parts that together transmit causal forces from X to Y. A mechanism describes each of the individual parts that together contribute to producing Y, disaggregated into entities engaging in activities (Machamer 2004). Measuring “quality of classroom instruction” is not enough to capture the causal mechanism; this is an intervening variable that tells us little about the actual causal process whereby vouchers increase the quality of classroom instruction, which then produces education attainment. According to Waldner, “Mechanisms explain the relationship between variables because they are not variables.” (Waldner forthcoming). If we believe we have measured the

mechanism by merely measuring an intervening variable, we are mistaken. The intervening variable can exist independent of a mechanism, and unless we actually engage in an in-depth case analysis tracing the mechanism itself using process tracing, we would not know whether there was a causal relationship linking X with Y through a mechanism involving quality of classroom instruction. Controlling for a “mechanism” measured as an intervening variable therefore does not allow us to control for a causal mechanism. This means that Gerring’s guidance makes little sense if mechanisms are actually taken seriously by the researcher.

Further, by downgrading mechanisms to mere intervening variables, Gerring’s book defines away the comparative strengths of process tracing methods, relegating the method to being a form of glorified storytelling. He describes process tracing as “thick description” on pp. 329–330. He then suggests that the “idea of process-tracing is also similar to judgments about context” (note 17, p. 306), and goes so far to say that process tracing has the function of being adjunct to larger analyses, making causal relationships more plausible (pp. 306–309), as we are “more likely to be convinced that X actually does (or does not) cause Y” when there is a narrative linked with it (p. 309).

Yet by investigating causal mechanisms using in-depth case studies, many scholars believe we are able to go a step deeper in our study of causal relationships (Bennett 2008a, 2010; Checkel 2008; Waldner 2010; Beach and Pedersen 2012). Even Gerring admitted in earlier work that if process tracing case studies are well constructed, they allow us to “peer into the box of causality to locate the intermediate factors lying between some structural cause and its purported effect” (Gerring 2007: 45). Note the use of the term “factors” instead of “variables,” and that he implies multiple factors instead of just an intervening variable that looks more like a mechanism.

Studying causal mechanisms with process tracing methods enables the researcher to make strong within-case inferences about the causal process whereby outcomes are produced, enabling us to update the level of confidence we have in the validity of a theorized causal mechanism (Beach and Pedersen 2012). The research process is, however, very demanding. If we are engaging in theory-testing process tracing, this requires that a causal mechanism is first conceptualized as a disaggregated series of parts composed of entities engaging in activities that together can hypothetically contribute to producing the outcome. The mechanism is then operationalized by developing case-specific empirical predictions of what evidence we should see in a case if each part of the causal mechanism is present. Empirical evidence is then collected to update our confidence in whether the hypothesized mechanism was present in the case.

Yet while being a very limited and demanding tool for studying causal mechanisms in single-case studies, process tracing methods have much stronger comparative advantages than Gerring admits. One would expect that in a book on social science methodology that intends to be useful to a majority of scholars, major methods such as process tracing would be better described, illustrating their comparative *strengths* and

uses instead of dogmatically focusing only on weaknesses relative to experiments.

Conclusions

While the book has many virtues, the overall verdict seen from a qualitative case-oriented researcher is one of disappointment and frustration. I simply do not understand Gerring’s ambition to create a unified framework in a manner that downplays the critical strengths and weaknesses of all different methodological approaches. In an ideal world, unity would be possible. But Gerring writes that “diversity at a foundational level is probably not helpful for the progress of social science... If causation means different things to different people then causal arguments cannot meet. Ships are allowed to pass silently in the night” (p. 376). However, one cannot through fiat define away these fundamental differences in ways that reduce other methods to being a “poor man’s substitute” for experiments. The result will end up alienating many qualitative, case study oriented scholars, just as KKV’s original book ended up doing.

As Gerring rightly states, methodology is all about trade-offs. Yet after reading the book, students will have the impression that there really is one good method (the experimental), which if applied correctly has few (if any) tradeoffs. All other methods are second/third-best alternatives with severe disadvantages relative to experiments. Is this really the understanding of social science methodology that we wish to teach our students? Or do we want them to be methodological pluralists, cognizant about *both* the strengths and weaknesses of different methods?

Note

¹ At an even more fundamental level, the book will be viewed with even more skepticism from qualitative scholars who adopt understandings of science that diverge from neopositivist/critical realist positions, such as what Jackson helpfully terms “analyticism”—a tradition where the main research goal is to *understand* the causes of *particular outcomes* (Jackson 2011). The “analyticism” understanding of social inquiry has become quite widespread amongst case-oriented researchers in recent years, best illustrated in Sil and Katzenstein’s (2010) recent edited volume on “eclectic theorization” that includes a number of prominent scholars presenting their own case-centric “eclectic” research designs. Although the authors in the book do not make their “analyticism” explicit, the idea of eclectic theorization is firmly situated within this tradition, given that it only makes sense to engage in “eclectic” theorization within the context of a single case study, where the goal is to account for the puzzling aspects of the particular case, yet following Gerring’s unified logic, “analyticism” is not good social science.

References

- Beach, Derek and Rasmus Brun Pedersen. 2012. *Process Tracing: Foundations and Guidelines*. Ann Arbor: University of Michigan Press.
- Bennett, Andrew. 2008a. “Process-Tracing: A Bayesian Perspective.” In *The Oxford Handbook of Political Methodology*. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, eds. (Oxford: Oxford University Press), 702–721.

- Bennett, Andrew. 2008b. "The Mother of all 'Isms': Organizing Political Science around Causal Mechanisms." In *Revitalizing Causality: Realism about Causality in Philosophy and Social Science*. Ruth Groff, ed. (London: Routledge), 205–219.
- Bennett, Andrew. 2010. "Process Tracing and Causal Inference." In *Rethinking Social Inquiry: Diverse Tools Shared Standards*, 2nd edition. Henry E. Brady and David Collier, eds. (Lanham, MD: Rowman & Littlefield), 207–220.
- Brady, Henry E. and David Collier, eds. 2010. *Rethinking Social Inquiry: Diverse Tools Shared Standards*, 2nd edition. Lanham, MD: Rowman & Littlefield.
- Bunge, Mario. 2004. "How Does It Work?: The Search for Explanatory Mechanisms." *Philosophy of the Social Sciences* 34:2, 182–210.
- Checkel, Jeffrey T. 2008. "Tracing Causal Mechanisms." *International Studies Review* 8:2, 362–370.
- Freedman, David A. 2010. "On Types of Scientific Inquiry: The Role of Qualitative Reasoning." In *Rethinking Social Inquiry: Diverse Tools Shared Standards*, 2nd edition. Henry E. Brady and David Collier, eds. (Lanham, MD: Rowman & Littlefield), 221–236.
- Gaines, Brian J., James H. Kuskliniski, and Paul J. Quirk. 2007. "The Logic of the Survey Experiment Reexamined." *Political Analysis* 15:1, 1–20.
- George, Alexander L. and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge: MIT Press.
- Gerring, John. 2007. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Gerring, John. 2010. "Causal Mechanisms: Yes But..." *Comparative Political Studies* 43:11, 1499–1526.
- Glennan, Stuart S. 1996. "Mechanisms and the Nature of Causation." *Erkenntnis* 44:1, 49–71.
- Glennan, Stuart S. 2005. "Modeling Mechanisms." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36:2, 443–464.
- Hall, Peter A. 2003. "Aligning Ontology and Methodology in Comparative Politics." In *Comparative Historical Analysis in the Social Sciences*. James Mahoney and Dietrich Rueschemeyer, eds. (Cambridge: Cambridge University Press), 373–406.
- Hedström, Peter and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36, 49–67.
- Jackson, Patrick T. 2011. *The Conduct of Inquiry in International Relations*. London: Routledge.
- Janis, Irving L. 1983. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Boston: Houghton Mifflin Company.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Machamer, Peter. 2004. "Activities and Causation: The Metaphysics and Epistemology of Mechanisms." *International Studies in the Philosophy of Science* 18:1, 27–39.
- Mahoney, James. 2001. "Beyond Correlational Analysis: Recent Innovations in Theory and Method." *Sociological Forum* 16:3, 575–593.
- Mahoney, James. 2007. "Qualitative Methodology and Comparative Politics." *Comparative Political Studies* 40:2, 122–144.
- Mahoney, James. 2008. "Toward a Unified Theory of Causality." *Comparative Political Studies* 41:4/5, 412–436.
- Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- Sil, Rudra and Peter J. Katzenstein. 2010. *Beyond Paradigms: Analytical Eclecticism in the Study of World Politics*. Basingstoke: Palgrave Macmillan.
- Waldner, David. 2010. "What are Mechanisms and What are They

- Good For?" *Qualitative and Multi-Method Research* 8:2, 30–34.
- Waldner, David. forthcoming. "Process Tracing and Causal Mechanisms." In *Oxford Handbook of the Philosophy of Social Science*. Harold Kincaid, ed. (New York: Oxford University Press)

Does Set-Relational Causation Fit into a Potential Outcomes Framework? An Assessment of Gerring's Proposal

Carsten Q. Schneider

Central European University, Budapest
schneiderc@ceu.hu

Ingo Rohlfing¹

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

Introduction

One of John Gerring's aims in his intriguing treatment of social science methodology is the development of a unified account for causal inference on the basis of the potential outcomes (PO) framework. Over the past two decades, the PO framework has become central in quantitative analyses (Morgan and Winship 2007). In qualitative research, in contrast, set theory and set-relational (SR) forms of causation and empirical research have started to play an ever more important role of unifying hitherto unrelated streams of qualitative literature (Goertz and Mahoney 2012; Ragin 1987, 2000, 2008; Rohlfing 2012: chap. 1; Schneider and Wagemann 2012). According to Gerring (2012: 337), the PO account is the more general framework and is able to accommodate SR causation such as necessity and sufficiency. In our contribution to this symposium, we discuss the viability of Gerring's proposal on how to perform SR research on the basis of the PO framework. We note here that the quest for common methodological ground in causal analyses is important and Gerring's suggestions in this regard are novel. Our critical reflections are therefore not meant to question the quest for a unifying framework *per se*.

In section two, we briefly introduce some basics of the PO and SR framework. In section three, we outline Gerring's two-step proposal of how to analyze set relations from a PO perspective. The fourth section includes a critical reflection of Gerring's proposal, arguing that it falls short in correctly capturing several core features of SR causation. Most importantly, we show that the suggested procedure can produce false negatives—indicating the absence of a set relation when, in fact, one exists—and false positives—suggesting the presence of a set relation when there is none. In the concluding section, we detail some of the most important features of SR causality. If the PO and SR frameworks are truly compatible, all of these SR features must be transposed into the PO framework.

The Potential Outcomes Framework and Set-Relational Causation—Some Basics

The discussion of differences and similarities between PO and SR is best achieved by presenting them in 2x2 tables. In the simplest version of PO, both the treatment (the cause) and the outcome are binary. Likewise, the simplest SR approaches relate a single crisp-set condition (the cause) to a crisp-set outcome.² In a PO framework, the combination of a binary treatment and outcome yields four different combinations (Table 1). The treatment can either be received (X) or not (~X), and one can or cannot observe the outcome for treated units (Y_X) and nontreated units (Y_{~X}).³ Two of these four scenarios can be observed, while the other two cannot (Morgan and Winship 2007: 35). We can observe the outcome for those units that were treated (cell A), and we can also observe the outcome for those units that were not treated (cell C). However, we cannot observe the outcome for treated units if they were not treated in fact (cell B), and the outcome for untreated units if they were treated (cell D). The four scenarios are used for the assessment of *treatment effects*, that is, the difference the presence and absence of the treatment makes for the outcome. For example, we can ask: What difference does it make for the success of a party in national elections when it launches negative campaign advertisement as opposed to a campaign that lacks negative advertisement?

Table 1: Potential Outcomes Framework

Outcome	For Treated (Y _X)	D	A
Observed	For Untreated (Y _{~X})	C	B
		No (~X)	Yes (X)
		Treatment Received	

The PO literature has developed numerous types of treatment effects, some of which are summarized by Gerring (2012: 223). Given the implicit focus of Gerring’s discussion, we limit ourselves to the *average treatment effect* (ATE).⁴ The ATE captures the difference between the average outcome of the treated for treated units (cell A) and the average outcome of the untreated for non-treated units (cell C). Since, in our example, the outcome is binary, inquiries into treatment effects ask for the likelihood of observing the outcome, given that units receive or do not receive the treatment. In formal terms:

$$ATE = p(Y_X|X) - p(Y_{\sim X}|\sim X)$$

With binary conditions and outcomes, p(Y_X|X) is equivalent to the ratio of cases in cell A, p(X ∩ Y_X), relative to all cases in the right column, p(X). Likewise, p(Y_{~X}|\sim X) is derived from the ratio of cases in cell C, p(\sim X ∩ \sim Y_{~X}), relative to all cases in the left column, p(\sim X). Clearly, then, the ATE relies on a *symmetric* notion of causation because a treatment makes a difference if there is a significant difference in the outcome between cases of X vis-à-vis those of \sim X.

The conventional understanding of SR causation, in contrast, is *asymmetric* (Ragin 1987; 2000; 2008; Schneider and Wagemann 2012) and thus diametrically opposed to the PO

framework. The difference can be seen once we ask which cells matter for the assessment of SR causation.

Table 2: Sufficiency and Necessity in 2x2 Tables

Outcome	Present	D	A
	Absent	C	B
		Absent	Present
		Condition	

The presence of a condition is *necessary* for the presence of the outcome if all cases that lack the condition are also non-members of the outcome. Formally, this means that p(\sim Y|\sim X) = 1.⁵ From an SR perspective, claims of necessity only generate expectations about cases that are not members of the condition, i.e., about cases in the left column of table 2 (Gerring 2012: 339).⁶ We expect cell D to be empty and cell C to contain cases. Similarly, the presence of a condition is *sufficient* if all cases that share the condition are also members of the outcome. Formally, it holds that p(Y|X) = 1. For sufficiency, only cases that share the condition are relevant, i.e., cell B should be devoid of cases and cell A should contain cases (Gerring 2012: 340). Finally, a condition is both *necessary and sufficient* if all members of the outcome set are also members of the condition set and vice versa. Formally: p(\sim Y|\sim X) = 1 ∩ p(Y|X) = 1. All that is required for a condition to be both necessary and sufficient is that cells B and D are devoid of cases and that cells A and C contain cases.⁷

An assessment of Gerring’s proposal on how to handle SR requires a short overview on how the latter are currently analyzed by SR scholars. In brief, SR researchers first assess whether the condition set is sufficient, necessary, or neither of the two. Those conditions that are either necessary or sufficient are assessed with regard to their *empirical importance*. For step one, the measure of *consistency* is employed, for step two that of *coverage* (Ragin 2006; Schneider and Wagemann 2012: chaps. 5, 9). Consistency measures the degree to which a distribution of cases is in accord with a postulated set relation. When interested in the sufficiency of X for Y, consistency is calculated by the share of cases in cell A relative to all cases in the right-hand column of the 2x2 table. The more cases that fall into cell B, the less consistent is the empirical evidence with the set-theoretic statement of sufficiency. In applied SR, deviations from perfect consistency are allowed. It follows that if consistency is higher than a specified threshold value, X is interpreted as sufficient. If consistency is lower, X is not considered as sufficient for Y.⁸ The *degree* of consistency, thus, is only used to differentiate sufficient from non-sufficient set relations and not to express the degree of importance of a sufficient condition.

Instead, it is coverage that expresses the empirical importance of a sufficient condition by asking: What is the share of cases that are members of the outcome and the condition relative to all cases displaying the outcome? This question is answered by count-ing the number of cases in cell A relative to all cases in the upper row. The lower the coverage, the lower is

the empirical importance of the sufficient condition because fewer cases are captured by the sufficient condition at hand. In contrast to consistency, there is no lower threshold for coverage below which conditions need to be dismissed as sufficient conditions.

Consistency has the same meaning in analyses of necessity. It expresses the degree to which the empirical evidence is in line with the statement that X is necessary for Y and is derived by the share of cases in cell C relative to all cases in the left column. Only conditions that pass a consistency threshold are interpreted as necessary conditions.⁹ Conditions deemed adequately consistent to be interpreted as necessary are subjected to an evaluation of their empirical relevance. The standard assessment involves calculating the share of cases that are members of both the condition and the outcome (cell A) relative to all cases that are members of the condition.¹⁰ As will become apparent below, consistency and coverage will play a central role in our assessment of Gerring’s proposal for a PO-centered analysis of set relations, which is introduced in the following section.

Gerring’s Procedure for a Potential Outcomes Analysis of Set Relations

Gerring (2012: 337–342) proposes a two-step approach for the analysis of set relations within a PO framework. In *step one*, the ATE is calculated to determine if X makes a difference to Y. For sufficiency, the rationale is that if X is fully consistent with a pattern of sufficiency, some cases are located in cell A while cell B is empty. Moreover, the expectation is that some cases are located in cells C and D. We therefore observe a difference in the distribution of cases in the left-hand and right-hand columns. Since the columnwise distributions can be captured as conditional probabilities, which, in turn, are the components of the ATE, a sufficient condition thus gives rise to an ATE that is different from zero. The reverse reasoning applies for necessity. If X is a fully consistent necessary condition, there should be cases in cell C and no cases in cell D. The absence of X makes a difference if cells A and B contain some cases, again creating a difference between the distribution of cases in the left and right columns and a nonzero ATE. However, a nonzero ATE alone does not suffice to infer *whether or not* X is a cause of Y, as this hinges on whether the ATE is *statistically significant* (Gerring 2012: 340).

Step two of the protocol aims at determining the *degree* to which X is necessary and sufficient, respectively. A significant ATE is an inadequate tool for this because its calculation draws on all four cells of the 2x2 table, whereas the analysis of any set relation only relies on two of the four cells. Since the same ATE can result from many different distributions of cases across the 2x2 table (see below), it is mandatory to look at how the cases are distributed across either the left or the right column. The left column captures the degree of necessity, expressed by the likelihood $p(\sim Y|\sim X)$. The closer it is to zero, the more the pattern is in line with a statement of necessity. Correspondingly, the degree of sufficiency is expressed by the likelihood $p(Y|X)$. The closer it is to one, the higher the degree of sufficiency (Gerring 2012: 340).

For an illustration of the two-step procedure utilized in an analysis of sufficiency, Table 3 reproduces an empirical example provided by Gerring (2012: 338–340). The ATE is calculated by computing the probabilities of observing Y under X and $\sim X$, respectively. These probabilities are 1, $p(Y|X)$, and 0.5, $p(Y|\sim X)$, yielding an ATE of $1 - 0.5 = 0.5$. Since the ATE is statistically significant with a t-score of about -22 , we proceed to step two. A look at the two columns shows that we are dealing with a condition that is fully in line with a pattern of sufficiency due to a conditional probability of $p(Y|X) = 1$. At the same time, X can hardly be considered a necessary condition because the likelihood of $p(Y|\sim X)$ is 0.5 and thus very low.

Table 3: ATE for Sufficiency for Gerring’s Hypothetical Example

Outcome	Present	250	500
	Absent	250	0
		Absent $p(Y \sim X)=0.5$	Present $p(Y X)=1$
Condition			

The example shows that the two-step approach is built on a division of labor. Step one answers “the question of causality” (Gerring 2012: 340) by differentiating between causes and non-causes. If the ATE is statistically significant, then X is a cause (if supplemented by causal mechanisms). Otherwise, it is not. Step two addresses the “question of probability” (Gerring 2012: 340) and then only asks for the degree to which X is sufficient and necessary, respectively. This implies four salient issues that are important to our attempt at evaluating the two-step proposal from an SR perspective.

First, the p-score of the ATE is the *only* basis for inferring causality (Gerring 2012: 340). If the ATE is significant, then the only unanswered question is how much X conforms to a specific form of set relation. If the ATE is insignificant, the analysis stops with the conclusion that there is no SR cause. Second, if a statistically significant ATE is detected, the columnwise conditional likelihoods only capture the degree to which we observe a set relation. Third, the two previous points imply that the *size* of the ATE is irrelevant. Both—whether X is an SR cause and to what extent—are captured by the p-score and a columnbound conditional likelihood. This implies that the *significance* of the ATE is the only PO framework element that carries some analytic weight in the analysis of SR causation. Fourth, the question of the empirical importance of an SR cause is left unaddressed by the two-step procedure as it exclusively focuses on statistical significance.

Five Critical Reflections at a Glance

Although compelling at first glance, we see five critical issues with the two-step procedure. Taking a broad view, the criticisms can be sorted into two rubrics. First, we question the claim that the PO framework can accommodate SR causation. Second, we cast doubt on the argument that the two-step protocol produces valid inferences on set relations.

The Ambiguity of the ATE: Illustrations

Less Unifying than it Seems

Our skepticism concerning the unifying potential of the PO account is based on four observations. First and quite obvious, the PO framework alone does not suffice for the analysis of set relations: it is integral to step one, but irrelevant to step two in which only ordinary conditional probabilities are calculated. Second and relatedly, one cannot distinguish between patterns of sufficiency or necessity based on the ATE alone. In the presence of a significant ATE, one must resort to column-wise conditional probabilities in order to determine the degree to which a set relation is present.

Third, step two of the procedure resembles the assessment of consistency in SR research. This is problematic because step two would reinvent a set-theoretic wheel and unnecessarily unsettles the semantic field (Gerring 2001: 39-40). More importantly, it inverts the sequence of analyzing set relations as practiced in SR research wherein consistency tests must go first because they carry the information on whether or not X can be considered as a SR cause at all. Instead, in the two-step procedure, the conditional probabilities are not used to decide whether X is a set relational cause, but to express the degree to which a set relation is given. The notion of degree of sufficiency suggests that there is no threshold probability below which X ceases to be a sufficient condition. The two-step procedure, thus, can lead to statements that the degree of sufficiency is, say, 0.2. This is nonsensical from an SR perspective because a condition with a consistency of 0.2 definitely cannot be considered an SR cause.

Fourth, we argue that empirical research should also make statements about the empirical importance of conditions, something the two-step procedure does not allow for but is easily accommodated in SR framework through the parameter of coverage.

PO and valid set-relational inferences

The fifth and perhaps most important critique is that the significance test of the ATE is uninformative about the presence or absence of set relations. To illustrate this for the case of sufficiency, we provide empirical examples showing that the ATE can be insignificant in the presence of a sufficient condition—a *false negative*—and that the ATE can be significant in the presence of a non-sufficient condition—a *false positive*. This is particularly troublesome because the significance of the ATE is the only element in the PO framework that assumes an inferential role in the assessment of set relations.

In order to claim that the significance of the ATE cannot correctly identify set relations, we need to rely on a measure that can do so. We deem the consistency score as a plausible measure because it is in line with the meaning of necessity and sufficiency as defined in set theory and formal logic. We think it is fair to argue that if the two-step procedure and the current SR best practice prompt different conclusions, then this casts doubts on the two-step procedure because its set-theoretic pedigree is less clear than that of consistency (and coverage).¹¹ In the following, we use hypothetical data and address these criticisms in more detail.

Significant ATE, Unclear Set Relation

Table 4 includes two distributions that produce an ATE of the same size and with the same p-score. However, the panels display two very different set relations, namely a fully consistent pattern of necessity in the upper panel and a fully consistent pattern of sufficiency in the lower panel.

Table 4: Same ATE, Different Set Relation

Outcome	Present	0	5
	Absent	10	5
		Absent $p(Y \sim X) = 0$	Present $p(Y X) = 0.5$
Condition			
ATE = 0.5; p = 0.01			

Outcome	Present	5	10
	Absent	5	0
		Absent $p(Y \sim X) = 0.5$	Present $p(Y X) = 1$
Condition			
ATE = 0.5; p = 0.01			

The significance of the ATE cannot discriminate between the two as it is insensitive to how the cases are distributed in the left and right columns of the panel. It is precisely for this reason that step two is required in Gerring’s protocol. The inspection of the column-wise conditional probabilities, however, is not an element of the PO framework. To us, this inability of the ATE to differentiate between necessity and sufficiency casts doubt on the more general claim that the PO framework can accommodate SR research.

We note that in the original exposition, step two of the protocol is only about the *degree* to which a set relation is present. This means that the two-step protocol does not require researchers to classify X as either a sufficient or a necessary condition. In this light, the problem that we identify here is not a problem because the upper and lower panels capture different degrees to which X is necessary and sufficient. However, we think that not classifying a condition as either necessary or sufficient or nothing is at odds with the established, qualitative view on set relations. While we agree that in applied empirical research it is warranted to allow for some deviation from perfectly consistent set relations, there nevertheless has to be a lower bound separating sufficient or necessary causes from non-SR causes. The argument that, given a significant ATE, X is always sufficient and necessary to some extent is very unusual.

Insignificant ATE and Consistent Set Relations

A further problematic issue is that the (in-)significance of

the ATE is a fallacious criterion for inferring the presence or absence of a set relation. For illustration, we use the identical distribution of a smaller number of cases, as in Gerring’s example above. Our example only includes 12 cases, but these cases are distributed across the cells so that they produce the same column-wise conditional probabilities and the same ATE. A significance test shows that this reduces the t-score of the ATE from about –22 in the original example (N = 1000) to –1.83 with a p-value of .10 (two-sided). If we follow the convention and take .05 as the threshold for statistical significance, we have to conclude that X does not qualify as being sufficient. However, a set-theoretic researcher would surely draw a different conclusion because the lower-right cell is devoid of cases and X is a fully consistent sufficient condition for Y (and empirically relevant with a coverage score of 0.5).

Table 5: Insignificant ATE with Fully Consistent Sufficiency

Outcome	Present	4	4
	Absent	4	0
		Absent $p(Y \sim X) = 0.5$	Present $p(Y X) = 1$
Condition			
ATE = 0.5; p = 0.10			

When one subordinates the traditional understanding of set relations under the PO framework, the conclusion that X is not sufficient is unproblematic, for it resolves the question of SR causality by exclusive reliance on the significance test of the ATE. However, for those who are, like us, convinced that it is a good idea to follow the definition of set relations anchored in formal logic and set theory, the example casts serious doubt on the suitability of the two-step framework for set-theoretic analyses.

The more general insight is that if we keep the column-wise conditional probabilities fixed, SR inferences derived from the two-step procedure depend entirely on the number of cases. This is of practical importance because empirical SR research often operates on small to medium-sized Ns (Rihoux et al. 2012). More precisely, the example shows that the two-step account can produce *false negatives*; the non-significant ATE makes us believe that X is not sufficient when, in fact, it clearly is from an SR perspective.

Significant ATE and Inconsistent Set Relations

The two-step procedure also suffers from the pitfall of *false positives*, meaning that the ATE is significant in the absence of any set relation from an SR perspective. The prospect of false positives stems from the inappropriate handling of set-relational inconsistency by the PO framework. The example in Table 6 yields a highly significant ATE of 0.40. However, the consistency of X as a sufficient condition is a dismal 0.60, far too low to interpret the condition as being sufficient according to standards in the current SR literature.

Table 6: Significant ATE with Inconsistent Set Relations

Outcome	Present	20	60
	Absent	80	40
		Absent $p(Y \sim X) = 0.2$	Present $p(Y X) = 0.6$
Condition			
ATE = 0.4; p < 0.01; consistency: 0.6			

Why do the significance of the ATE and consistency convey different messages in this example? Apart from the large number of cases, which always boosts statistical significance, the difference stems from the way in which ATE and consistency treat the distribution of cases in the left column ($\sim X$). Consistency, staying true to the asymmetric nature of set relations, does not take the distribution in the left-hand column into account at all. The significance of the ATE, in contrast, also depends on how instances of $\sim X$ are distributed. In table 6, the ATE is significant because most instances of $\sim X$ are also cases of $\sim Y$ (80 out of 100). This means that the ATE is crucially shaped by a relatively large number of cases in the lower left cell. Thus, the inference that X is sufficient of Y is (largely) driven by cases that are neither members of X nor of Y.¹² From an SR perspective, this is a misguided procedure and inference. One might reject this criticism by pointing out that step two of the procedure only establishes the degree to which X is sufficient and necessary, respectively. In our empirical example, this would yield the statement that X is sufficient to a degree of 0.6 and necessary to a degree of 0.2. This, in our eyes, amounts to a nonsensical interpretation of set relations, though.

PO and Empirical Importance of Conditions

The two-step procedure conveys information on whether a set relation is given and the degree of sufficiency and necessity. What is missing from the procedure is an assessment of empirical importance. In light of what we have said about the three elements of the PO framework above, the only quantity that serves no other purpose and could act as the measure of empirical importance is the size of the ATE. The larger the ATE, the more important the condition would be. However, the example in Table 7 shows that the *size* of the ATE is not a valid measure for substantive importance in SR research as a *case-based* approach (Ragin 1987). With a size of 0.80, the ATE is relatively large and highly significant, which, in this example, matches a consistency score of 1.¹³ However, the coverage of the sufficient condition X is very small with a score of 0.10.

The difference between the size of the ATE and coverage is again due to the large number of cases in the lower-left cell. *Ceteris paribus*, the more cases that are located in this cell, the larger the size and significance of the ATE. In contrast, the coverage score remains entirely unaffected because its assessment only depends on cases in the upper row. SR research focuses on the upper row because it includes the cases that display the outcome, which is what we are interested in and seek to explain in SR analyses (Ragin 2008: chap. 11).

Table 7: Large ATE, Small Coverage for Sufficiency

Outcome	Present	90	10
	Absent	450	0
		Absent $p(Y \sim X) = 0.2$	Present $p(Y X) = 1$
Condition			
ATE = 0.8; $p < 0.01$; coverage: 0.1			

Summary

The two-step approach attempts to reconcile the asymmetry of SR causation with the symmetric notion of causation inherent to the PO framework. The critical issues that we raised above suggest that the PO and SR frameworks do not fit easily together, thereby reinforcing conclusions previously reached by other scholars in different contexts (see Braumoeller and Goertz 2000; Clarke 2002; Mahoney 2008; Seawright 2002a, b).

One major reason for this misfit is that the first of the two-step procedure draws on cases from *all* cells of the 2x2 table, while set theory tells us that only one of the two columns is relevant for SR inferences. From a set-theoretic perspective, only the right column is of direct relevance for sufficiency, and only the left column for necessity. We do not claim that the two-step procedure and reliance on all four cells *always* produces false positives or false negatives. However, whether or not the two-step procedure and established SR analyses lead to the same conclusion (as it does in Gerring’s example, 2012: 339–340) depends on the distribution of the data and is therefore accidental.

Instead of a Conclusion—Where To Go From Here?

In concluding our contribution to the symposium, we deem it worthwhile to at least briefly address some of the hitherto unmentioned additional challenges in integrating the SR and PO framework. First, all examples discussed here and in Gerring (2012: 339) are drastically reductionist, for they assume that SR researchers are interested in single conditions. In practice, however, researchers invoking SR causation usually do not do so because they are interested in equifinality, unifinality, conjunctural causation, and INUS and SUIN conditions (Goertz and Mahoney 2012; Rohlfing 2012: chap. 2; Schneider and Wagemann 2012). This focus on *causal complexity* triggers a series of complications that cannot be discussed in detail here. For instance, the SR focus on conjunctions of conditions as opposed to single conditions cannot be adequately captured by interactions. Classic statistical analysis of interactions becomes overly complicated with more than three variables (Braumoeller 2004), whereas SR routinely manages higher-order conjunctions.¹⁴ In contrast to Gerring (2012: 356), we also think that the factorial design does not capture the same essence as SR conjunctions for reasons too long to be explored here.

Second, those who attempt to unify PO and SR must also come to terms with the fact that the latter often operates in a *small-N to medium-N setting*, thus raising doubts about the

usefulness of statistical significance as a means for detecting set relations. By this, we do not mean that notions of statistical significance cannot, or should not, be applied to set relations. The SR literature has made several suggestions in this regard (Dion 1998; Eliason and Stryker 2009; Ragin 2000). These tests, however, must be designed such that they only take into account those cases that matter from a set-theoretic perspective (Braumoeller and Goertz 2000). Third, and relatedly, unifying PO and SR requires keeping separate rather than conflating or neglecting the set-theoretically grounded parameters of fit *consistency and coverage*. Put differently, we need clarity with regard to the PO status of consistency and coverage. Fourth, advanced SR has, of course, long moved beyond dichotomous sets (crisp sets), and has embraced the notion of *fuzzy sets* (Ragin 2000). Since fuzzy sets are distinct from continuous variables (Ragin 2008: chap. 4), the PO literature on continuous treatments is only of limited use when trying to gauge how PO and fuzzy sets relate to each other.

We conclude that PO and SR frameworks might be compatible and that it is absolutely worthwhile to find ways in which they can be integrated. However, we also think that unless it can be shown that all core notions of SR causation can be adequately managed within the PO framework, it might be wiser to work under the assumption that the two are not reconcilable.

Notes

¹Rotation principle. Both authors contributed equally to the paper. We thank Achim Goerres for helpful comments.

²The PO and SR framework can work with more than one condition and these conditions and the outcome can be continuous (PO) and fuzzy sets (SR), respectively. In addition, SR subsumes notions of equifinality, conjunctural causation, and INUS and SUIN conditions. We postpone a discussion of some of these issues to a later section. For the time being, we follow Gerring in juxtaposing the basic setup.

³In order to achieve some consistency between the notation in the PO framework and SR analyses, we use $\sim X$ to denote the absence of the treatment (\sim meaning “not”).

⁴Most of our arguments also apply to other treatment effects.

⁵We drop subscripts to Y in the following, keeping in mind that we are discussing the ATE.

⁶We follow Gerring in using this classic definition from logic. An alternative and more common definition in the methods literature is $p(X|Y) = 1$ (Goertz and Mahoney 2012). Both are equivalent and our arguments are independent of the definition used.

⁷Conditions that are both necessary and sufficient are rare in social science theory and empirical research alike. We therefore do not discuss them any further here.

⁸The precise consistency threshold varies with features of the analysis at hand (Schneider and Wagemann 2012: chap. 5), but should never be lower than 0.75 for sufficiency (Ragin 2006).

⁹For necessary conditions, the minimum consistency value should not be lower than 0.9 (Schneider and Wagemann 2012: chap. 9).

¹⁰Goertz (2006) suggests a slightly different procedure, whereas Schneider and Wagemann (2012) propose a formula for assessing the relevance of a necessary condition that integrates and elaborates on both Goertz’ and Ragin’s (2006) proposal.

¹¹This still leaves room for the conclusion that SR causes are not causes at all but simply descriptive statements about the social world

(King and Powell 2008).

¹² We do not have a pattern of necessity here because consistency for necessity is only 0.80.

¹³ Gerring (2012: 338) issues the caveat that a cause should not be trivial. To us, it is not entirely clear what trivial means here. It seems that X counts as trivial when the column-wise conditional likelihoods are identical because then X does not make a difference to Y. However, it is not apparent why this should designate a trivial cause, as similar conditional likelihoods simply denote that X is a non-cause in a PO perspective. It is questionable to exclude all scenarios in which X does not make a difference as trivial, as this arbitrarily limits the application of the two-step procedure to studies about which we know that X makes a difference. In any case, it is safe to argue that Gerring's understanding of trivialness is different from the established SR understanding. In SR research, a condition is trivial when its coverage is negligible.

¹⁴ Braumoeller's (2003) Boolean probit and logit seem more promising as they accommodate high-order interactions and equifinality. However, apart from the enormous data requirements in terms of the numbers of cases, one has to specify the conjunctions and the substitutability of conjunctions (a.k.a., equifinality) in advance of the empirical analysis. One might consider this beneficial from the viewpoint of hypothesis testing, but it increases the risk of model misspecification. This is different for SR research (and *Qualitative Comparative Analysis* in particular), allowing one to test specific models and realize more exploratory analyses deriving equifinal conjunctions (or single conditions) from a menu of conditions.

References

- Braumoeller, Bear F. and Gary Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science* 44:4, 844–858.
- Braumoeller, Bear F. 2003. "Causal Complexity and the Study of Politics." *Political Analysis* 11:3, 209–233.
- Braumoeller, Bear F. 2004. "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization* 58:4, 807–820.
- Clarke, Kevin A. 2002. "The Reverend and the Ravens: Comment on Seawright." *Political Analysis* 10:2, 194–197.
- Dion, Douglas. 1998. "Evidence and Inference in the Comparative Case Study." *Comparative Politics* 30:2, 127–145.
- Eliason, Scott R. and Robin Stryker. 2009. "Goodness-of-Fit Tests and Descriptive Measures in Fuzzy-Set Analysis." *Sociological Methods Research* 38:1, 102–146.
- Gerring, John. 2001. *Social Science Methodology: A Criterial Framework*. Cambridge: Cambridge University Press.
- Gerring, John. 2012. *Social Science Methodology: A Unified Framework*. Cambridge: Cambridge University Press.
- Goertz, Gary. 2006. "Assessing the Trivialness, Relevance, and Relative Importance of Necessary and Sufficient Conditions in Social Science." *Studies in Comparative International Development* 41:2, 88–109.
- Goertz, Gary and James Mahoney. 2012. *A Tale of Two Cultures: Contrasting Qualitative and Quantitative Paradigms*. Princeton: Princeton University Press.
- King, Gary and Eleanor Neff Powell. 2008. "How Not to Lie without Statistics." Paper delivered at the Annual Meeting of the American Political Science Association.
- Mahoney, James. 2008. "Toward a Unified Theory of Causality." *Comparative Political Studies* 41:4–5, 412–436.
- Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactuals*

and Causal Inference: Methods and Principles for Social Research. New York: Cambridge University Press.

- Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Quantitative and Qualitative Strategies*. Berkeley: University of Berkeley Press.
- Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- Ragin, Charles C. 2006. "Set Relations in Social Research: Evaluating Their Consistency and Coverage." *Political Analysis* 14:3, 291–310.
- Ragin, Charles C. 2008. *Redesigning Social Inquiry*. Chicago: Chicago University Press.
- Rihoux, Benoît, Priscilla Álamos, Damien Bol, Axel Marx and Ilona Rezsöhasy. 2012. "From Niche to Mainstream Method? A Comprehensive Mapping of QCA Applications in Journal Articles from 1984 to 2011." Manuscript under review.
- Rohlfing, Ingo. 2012. *Case Studies and Causal Inference: An Integrative Framework*. Basingstoke: Palgrave Macmillan.
- Schneider, Carsten Q. and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences. A Guide to Qualitative Comparative Analysis*. Cambridge: Cambridge University Press.
- Seawright, Jason. 2002a. "Testing for Necessary and/or Sufficient Conditions: Which Cases Are Relevant?" *Political Analysis* 10, 178–193.
- Seawright, Jason. 2002b. "What Counts as Evidence? Reply." *Political Analysis* 10:2, 204–207.

Concepts and Description in Gerring's Social Science Methodology, 2nd ed.

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

In the Jorge Luis Borges' short story "Funes, the Memorious," the eponymous Funes is blessed with astounding powers of recall and observation. So powerful is his memory that he insists on creating a new language, with unique signifiers for every aspect of existence:

Not only was it difficult for him to see [that] the generic symbol 'dog' took in all dissimilar individuals of all shapes and sizes, it irritated him that 'dog' of three-fourteen in the afternoon, seen in profile, should be indicated by the same noun as the dog of three-fifteen, seen frontally.

Ultimately, Borges' narrator concludes, Funes was "not very capable of thought. To think is to forget a difference, to generalize, to abstract. In the overly replete world of Funes there was nothing but details, almost contiguous details" (Borges 2000).

Like Funes, many social scientists, particularly qualitative-oriented researchers, talk about concepts all the time. But few spend much time thinking about them. A typical monograph devotes a paragraph or so in the introductions to "define" their key terms, justify the definitions by their careful balance of fidelity and alteration to a more senior authority in the field, and then move on to the task of detailing a historical narrative. Beyond such ritual recitations, there is a general

skepticism about the necessity for concepts. Robert Solow, for one, warns that while theories—and here we might easily add concepts as well—are “cheap, data are expensive” (Solow 1997). Karl Popper cautions against being “goaded into taking seriously problems about words and their meaning” (cited in Gerring 2012: 113). To some extent, this aversion to concepts must be forgiven in a discipline that prides itself on empiricism. Modern science rebels against the Platonic notion that human experience is limited only to a refracted and distorted perception of the pristine, abstract, and ultimately unobservable reality. Concepts are thus relegated to a branch of analytical philosophy, far divorced from the social science. That the entirety of Part II of the second edition of John Gerring’s *Social Science Methodology* (SSM2) is devoted to issues of conceptualization and measurement (about a quarter of the book overall) is a testament to the importance Gerring vests in the subject. In SSM2, Gerring adds two major themes to the study of concepts: First, in placing the discussion of concepts squarely in the contexts of methodology, SSM2 lays out comprehensive and explicit criteria for evaluating concepts in light of their specific role in causal analysis. Second, it develops an inductive complement to the predominantly deductive approaches to conceptualization and offers techniques of description as the link between abstraction and empirics.

The Ladder of Abstraction and Beyond

SSM2 arrives during what could be described as a burgeoning renaissance for the study of concepts. Unlike many books of its genre, the first edition of SSM in 2001 did not shirk from exploring conceptualization. Following that, Gary Goertz’s highly acclaimed *Social Science Concepts: A User’s Guide* (Goertz 2006) makes a strong case for the centrality of concepts to causal analysis, to name just one more prominent example. Still, it is hard to overlook the continued centrality of Giovanni Sartori to any discussion of concepts in the social sciences (Sartori 1970, 1984, 1993). Sartori’s “Concept Misformation in Comparative Politics” from 1970 is being cited at an ever increasing clip, with nearly half of the 276 total citations recorded in the ISI Web of Science databases occurring since 2005. In this and latter work Sartori broadsides common contemporary uses of concepts and offers his own guidelines for the proper use of concepts. Key to his argument is the metaphorical ladder of abstraction, which posits an inverse relationship between a concept’s *intension* (connotation), the systematic and explicit definition of the characteristics of the concept, and *extension* (denotation), the range of cases that can be categorized as meeting the conceptual definition. Moving up the metaphorical ladder of abstraction necessarily means losing specific definitional characteristics while expanding the number of cases to which a concept applies. Moving down the ladder means adding definitional specificity while shrinking the number of qualifying cases. This essentially Aristotelian approach to concepts using necessary and sufficient conditionality has been enormously influential in the study of regime type, creating new interest in typological hierarchies (Collier and Levitsky 1997; Møller and Skaaning 2010). The ladder of abstraction has also become a vehicle to explore

the differences between qualitative and quantitative methods, as the former is deemed to work with concepts that are thicker but narrower, the latter to work with concepts that are thinner but broader (Coppedge 1999).

Still, Sartori provides at best incomplete guidance for evaluating concepts. This is in part because the very notion of conceptual stretching is ironically imprecise. Sartori warns that concepts are often “stretched” or “strained” as researchers ascend or descend the ladder of abstraction. “In order to obtain a world-wide applicability, the extension of our concepts has been broadened by obfuscating their connotation.” Utilizing such concepts means research “cover[s] more—in traveling terms—only by saying less, and by saying less in a far less precise manner” (Sartori 1970: 1035). David Collier and James Mahon summarize conceptual stretching as the “the distortion that occurs when a concept does not fit new cases” (Collier and Mahon 1993: 845). But where does revising and amending conceptual definitions end and distortion begin? More importantly, if, as Sartori himself argues, moving between levels of abstraction is intrinsic to theory building, then how can we avoid such distortion?

Without clear answers from Sartori, much of the work on social science methodology has proceeded by avoiding discussion of concepts in general. A case in point is Gary King, Robert Keohane, and Sidney Verba’s highly influential *Designing Social Inquiry* (DSI) (King, et al. 1994). DSI contains only one substantive piece of advice regarding conceptualization—“maximize concreteness”:

The standard for explanation in any *empirical* science like ours must be *empirical* verification or falsification. Attempting to find empirical evidence of abstract, unmeasurable, and unobservable concepts will necessarily prove more difficult and less successful than for many imperfectly conceived specific and concrete concepts. (King et al., 1994: 110, emphasis in original)¹

It is hard to deny that this sounds sensible. But it also can be construed as a license for conceptual stretching, in essence advocating the use of a variable based on the availability of observable data. DSI admits that prioritizing “concreteness” (i.e., operationalization) contributes to a gap between concept and indicator, such as when the average length of employment within a federal agency is taken as a measure of “institutionalization,” amounting to a diminishment in the concept’s domain. Moreover, seeking to minimize abstraction (maximize concreteness) also pushes a concept to the lower rungs of the ladder of abstraction, where there are fewer and fewer empirical referents. DSI concedes that some measure of abstraction is necessary to “expand our frame of reference and the applicability of our theory” as well as to expand the case universe (another one of DSI’s cardinal methodological prescriptions) (King et al., 1994: 111).

In his discussion of concepts, Gerring moves away from such all-or-nothing propositions. What really matters in evaluating concept formation is when, where, and in what ways the concepts themselves are being used. Following the overall theme that every methodological venture involves trade-offs

and the goal should be to “to maximize a particular set of virtues while downplaying a corresponding set of vices” (381), Gerring offers a broader set of criteria for evaluating concepts than either Sartori or DSI considered:

1. **Resonance:** How faithful is the concept to the extant definitions and established usages?
2. **Domain:** How clear and logical is the relationship of a concept to the language community and the empirical terrain of cases?
3. **Consistency:** Is the meaning of the concept consistent within a single piece of research?
4. **Fecundity:** How many attributes do referents of a concept share?
5. **Differentiation:** How differentiated is a concept from neighboring concepts? What is the contrast-space against which a concept defines itself?
6. **Causal utility:** What utility does a concept have within a causal theory and research design?
7. **Operationalization:** How do we know it when we see it? Can a concept be measured without bias?

Cast in this broader frame, we see both the strengths and the limitations of both Sartori and DSI. Sartori’s guidelines, for instance, are subsumed largely in the notion of resonance, domain, internal consistency, and differentiation. In a faint echo of Funes, Sartori even inveighs against “terminological waste” and holds that “awaiting contrary proof, no word should be used as a synonym for another word” (Sartori 1984: 39). These are no doubt valuable suggestions, but also they are also limiting to anyone who is trying to use concepts in relationship to empirical observation and in causal accounts. At the same time, DSI’s insistence on “concreteness” seems to prioritize operationalization at all costs, with little regard for resonance, domain, or even consistency.

Gerring’s criterial approach allows greater flexibility to evaluate concepts based on the particular role they play in research. A project that begins as a game theoretical model might use relatively abstract conceptual definitions (“information” or “power”) that need only be malleable through algebra. First and foremost, as Sartori notes, the definitions need to have clear and logical scopes and consistent usages. The question of identifying empirical indicators—expanding the fecundity, differentiation, causal utility, and operationalization of the concept—can be secondary. On the other hand, a project that is data-driven, that begins by presenting rows of statistical data or deep ethnographic analysis, may generate extremely rich conceptual language before considering issues of resonance, domain, or even consistency. The next logical question, then, is how to improve concepts themselves.

Conceptualization as a Fishing Trip

One of the unfortunate legacies of being so beholden to Sartori for so long has been the sense that conceptualization itself is solely deductive, an abstract endeavor divorced from empirics. Sartori tended to strike a dismissive tone when considering the vagaries of actual observation as opposed to abstract reasoning; concepts can—dare we say, must—be

mapped on the ladder of abstraction by logical deduction alone without being sullied by the “growing potpourri of disparate, non-cumulative—and in aggregate—misleading morass of information.” Sartori abhors fishing for data “without adequate nets” of pre-existing conceptual categories (Sartori 1970: 1939). Unspecified “preliminary siftings”—actual observation—is relegated to an initial step that receives almost no attention (Sartori 1984: 47–49).

A number of scholars have sought to strike a better balance between inductive and deductive components in concept formation. DSI’s scant discussion of concepts, for instance, concludes that “if we have no alternative to using unobservable constructs, as is usually the case in the social sciences, then we should at least *choose ideas with observable consequences*” (King, et al., 1994: 111, emphasis in original). Robert Adcock and David Collier posit a kind of double helix in which “scoring” of cases can lead a researcher to refine indicators, and refining indicators leads to new scoring (Adcock and Collier 2001). Charles Ragin and other advocates of fuzzy-set analysis similarly emphasize that a recursive relationship between observations and concepts is necessary for effective calibration (Ragin 2000). Perhaps most significantly, Goertz shows that changes at the indicator level that affect how a concept is defined influence the universe of referent cases and in turn the causal relationship to be explained, thus making concept formation itself critical to causal analysis (Goertz 2006).

Gerring’s contribution in this regard, though, is to go even further in highlighting the empirical, inductive elements in concept formation. This element is wholly novel to SSM2 and represents a significant improvement over the previous edition. Gerring does not discount the importance of deduction and logical construal of conceptual domains, but the real value added is his attempt to bridge the empirical and the abstract by scrutinizing what data actually look like in comparison to the conceptual language. Without prejudicing the importance of the proper conceptual equipment by which data are accumulated, Gerring does not toss back the fish just because it was caught in an imperfect net. In Chapter 6, Gerring defends the practice of “mere description” as a critical and underdeveloped task in the social sciences. As he expounds, the very phenomena social scientists most want to explain—democracy, human rights, war, revolution, standards of living, mortality, ethnic conflict, happiness/utility, and inequality—remain generally under-described and therefore under-conceptualized. Rushing to offer etiological accounts about them, then, is the equivalent of trying to run before learning to walk.

Gerring catalogs four types of descriptive arguments, each associated with different and increasingly complex conceptual frameworks. We can consider each in turn: *Indication* focuses on individual features of cases without making any attempt to link those features. This corresponds to a relatively simple conceptual definition, or at least a basic configuration of proposed indicators. We may be interested, for instance, in the ages of voters or whether or not a country is democratic. With *syntheses* we begin to cluster a number of indicators around a single, multi-dimensional conceptual rubric. Here we might find the concept of advanced industrialized democracies, with

multiple indicators (wealth, industrial output, democracy, etc.) used to denote classification within a category. If synthesis is about lumping, then *typology* is the function of splitting (Zerubavel 1996). Typology differentiates concepts from one another, resolving synthetic clusters in separate groupings. Some concepts may share some common features but be different in others. For example, if the definition of advanced industrialized democracy is set, what can we make of countries that are democratic but not industrial or are industrial but are not democratic? Finally, *associations* move beyond categorization of cases and differentiation of concepts to identification of patterns based on location in specific spatial, temporal, or other types of networks. We could observe, for instance, that strong party systems exist where there is a first-past-the-post single-district electoral system or where democratization follows industrialization.

From a concept-builder's perspective, it is interesting to note the role that these descriptive acts have in aspects of concept formation. Sartori's main preoccupations are with resonance, domain, and consistency, elements that can be improved by *a priori* reasoning. Yet other components of Gerring's criteria, like fecundity, differentiation, and causal utility, depend on *a posteriori* observation. Moving from observation of discrete indicators to synthesis contributes to bolstering a concept's fecundity and increasing the number of individual attributes referents share. Differentiation of concepts occurs through typology, in which a concept is compared to its closest synonym and/or its polar opposite. While this might involve some element of deduction, it is also clearly informed by the tension between clustering and typologizing (i.e., separating) groups of observations. It also shows a potential trade-off between fecundity and differentiation. Finally, observation of correlative association is necessary but not sufficient for any claim of causal utility. Put differently, simply observing the presence of X and Y together does not prove causation, but demonstrating the absence of either X or Y would undermine any claim that there is a causal relationship between the two.

In order to demonstrate the power of this classification system, it is useful to shift from examining the social sciences, with its relative poverty of description, to medicine, where habits of data-collection are far more robust. The routine physical examination, from its collection of self-reported data (age, medical history, place of residence), to simple mechanical measurement (weight, height), to more complex chemical analysis (blood tests, etc.) yields far more data than most political scientists could ever hope to manage. Consequently, medical history is often changed by those who take the first step of understanding illness simply by observing and describing its manifestations.²

AIDS represents an especially interesting case because it shows the interaction between observation and disease conceptualization. The first observation of the disease noted the sudden appearance of a handful of gay men with similar symptoms of immune system collapse, such as Pneumocystis pneumonia and Kaposi sarcoma, a rare form of cancer. Based on weekly reports from the Centers for Disease Control (CDC), the early definitions of the disease incorporated this social

location within the definition itself, dubbing the disease variously as gay plague or gay cancer. Positing an association between the emergence of the disease in the gay community and lifestyle unique to that community, early searches for a cause focused on side-effects or impurities in illicit drugs commonly used in the homosexual community, complications with bacterial, viral, and parasitic infections common among practitioners of anal sex, "overexposure" to semen, or even water contamination in gay neighborhoods. But, says medical historian Jonathan Engel,

none of these explanations accounted for the nonhomosexuals who were becoming infected, or the hemophiliacs, or the Haitians. The explanation became even less satisfying when the CDC reported in late 1982 that at least 26 infants and children had also been diagnosed. In the following months investigators began to agree that the cause was most likely a blood-borne pathogen that was frequently [but not exclusively] transmitted sexually. (Engel 2006: 8)

This came about as researchers began to pool their empirical observations and recognized that a number of heterosexuals—who by definition could not have had the "gay" disease—nonetheless exhibited a cluster of similar symptoms of immune system collapse. Alighting on the term Acquired Immune Deficiency Syndrome represented an adjustment in the attributes and indicators of the concept as well as the referents, the population that could (or did) have the disease. Of course, the modification of the disease concept had no impact whatsoever on etiology. Yet it had a significant impact on the subsequent search for causes. Observation of discrete indications (appearance of unique and rare symptoms in a handful of patients) led to identification of clusters of symptoms (immune system collapse, etc.) in specific social groupings (initially gay men, then hemophiliacs and other population segments). This synthesis formed the basis of a new, more generalized, disease concept. It was still several years, though, until the HIV retrovirus was isolated in a single AIDS patient and differentiated (typologized) from other similar retroviruses, much less demonstrated to be present in the wider population of AIDS patients (and correspondingly absent in those without the disease), establishing the basis for an associational claim (Epstein 1996; Kallings 2008).

The social sciences lack a set of core concepts which can guide data collection comparable to medicine's, nor do they have an institutional infrastructure that can pool data like the CDC or National Institutes of Health. That does not mean, however, that simple, largely unguided acts of observation—mere fishing expeditions—have no role in the social sciences. In an interview regarding his pioneering research on authoritarian and sultanistic regimes, Juan Linz describes his dissatisfaction with the conceptual binary of democracy/totalitarianism that had dominated comparative politics in the 1950s and 1960s. Linz observed that regimes such as Franco's Spain, Trujillo's Dominican Republic, and Somoza's Nicaragua differed drastically from existing models of totalitarian Germany or the Soviet Union (Linz 2000). "First of all, I want to describe

reality,” he said.

There is a world out there, and I want to describe it in some way, just as a journalist or a historian does. But I want to describe the reality with a conceptualization more abstract than just telling a story. So, I try to describe reality, and then to conceptualize it. (Munck and Snyder 2007: 182)

It was only after the practice of descriptive argument—indication, synthesis, and typologizing—that Linz eventually grounding his conceptual terminology in Max Weber’s elaboration of the concepts of patrimonialism and sultanism.

Of course, conceptualization can proceed deductively, using existing concepts as a kind of raw material, as Dahl describes in elaboration of concept of polyarchy (See Munck and Snyder 2007: 125–128). And Sartori is correct in warning that unguided empirical searches often yield concepts that lack the potential for cumulation, are unmoored from existing uses, or do violence to older definitions. But there are gains that also accrue just from putting our nets in the water. Highlighting the inductive underpinnings of concept formation and thereby linking observation, description, and abstraction is one of SSM2’s singular achievements.

Conclusion

The danger of focusing too intently on methodological trade-offs is that it might tempt us to forget about “trade-ups,” the possibility of achieving a net or overall improvement. Gerring rejects the idea of unmitigated pluralism, concluding that “within the context of a given study some choices may be superior to others. That is, one set of concepts, arguments, and analyses may lead to a better reconciliation of methodological demands than another” (SSM2 382). As it pertains to conceptualization and description, though, we could justifiably ask for more guidance as to where the trade-ups are. Many have suggested that the recent trend of multimethod research can resolve the seemingly ubiquitous trade-offs between depth versus breadth. Coppedge, for one, speaks of “thickening thin” quantitatively-construed concepts through case study research nested in large-n analysis (Coppedge 1999; See also Lieberman 2005). Gerring cautiously endorses multimethod research, but does not elaborate upon its implications for concept formation. Yet there may still be no satisfactory way to overcome the obstinate Ladder of Abstraction. Rather than adding heft to thin concepts, we may devolve to the lowest common denominator by shredding the thickest of our concepts in order to fit them in the rows and columns of a spreadsheet (Ahran 2011).

Instead of thinking about SSM2’s contribution to the issues of conceptualization as a guide to overall improvement, we could think of it more as a field guide for the myriad ways different kinds of concepts emerge and evolve. Even in a discipline as replete with data as medicine, surgeon Atul Gawande urges all doctors to incorporate counting into their daily routines: How often are sponges accidentally left in after surgery? How long do patients wait in the lobby? More than that, he argues that every doctor should write something (a blog, a diary, or a journal article) about their practice daily, above and

beyond what is already incorporated in the standard patient history (Gawande 2007). In effect, doctors should try to be ethnographers of their own field, constantly seeking to identify and analyze patterns at an individual level. Perhaps Gerring’s most important contribution, then, is to show how the adoption of similar practices in the social sciences can help not only in collecting data but in building more robust concepts by which to organize them.

Notes

¹ Interestingly, Van Evera (1997: 47–48) seems to dissent from *DSI*, saying that rather than selecting concepts that “are easy to measure” we should to “give bonus credit to scholars who take on the harder task of studying the less observable.” He gives no indication of how to do that, however.

² For an interesting discussion about Robert Farber’s initial work on leukemia, see Mukherjee (2010). On the identification of anthrax, see Jones (2010).

References

- Adcock, Robert, and David Collier. 2001. “Measurement Validity: A Shared Standard for Qualitative and Quantitative Research.” *American Political Science Review* 95:3, 529–546.
- Ahran, Ariel. 2011. “Concepts and Measurement in Multimethod Research.” *Political Research Quarterly* (December).
- Borges, Jorge Luis. 2000. *Labyrinths*. J. Irby, trans. New York: Penguin Classics.
- Collier, David and James E. Mahon. 1993. “Conceptual Stretching Revisited: Adapting Categories in Comparative Analysis.” *American Political Science Review* 87:4, 845–855.
- Collier, David and Steven Levitsky. 1997. “Democracy With Adjectives: Conceptual Innovation in Comparative Research.” *World Politics* 49:3, 430–451.
- Coppedge, Michael. 1999. “Thickening Thin Concepts and Theories: Combining Large N and Small in Comparative Politics.” *Comparative Politics* 31:4, 465–476.
- Engel, Jonathan. 2006. *The Epidemic: A Global History of AIDS*. Washington DC: Smithsonian.
- Epstein, Steven. 1996. *Impure Science: AIDS, Activism, and the Politics of Knowledge*. Berkeley: University of California Press.
- Gawande, Atul. 2007. *Better: A Surgeon’s Notes on Performance*. New York: Metropolitan Books.
- Gerring, John. 2012. *Social Science Methodology: A Unified Framework*. Cambridge: Cambridge University Press.
- Goertz, Gary. 2006. *Social Science Concepts: A User’s Guide*. Princeton: Princeton University Press.
- Jones, Susan. 2010. *Death in Small Packages: A Short History of Anthrax*. Baltimore: Johns Hopkins University Press.
- Kallings, L.O. 2008. “The First Postmodern Pandemic: 25 Years of HIV/AIDS.” *Journal of Internal Medicine* 263:3, 218–243.
- King, Gary, Robert 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, 435–452.
- Linz, Juan. 2000. *Totalitarian and Authoritarian Regimes*. Boulder: Lynne Rienner.
- Møller, Jørgen and Svend-Erik Skaaning. 2010. “Beyond the Radial Delusion: Conceptualizing and Measuring Democracy and Non-Democracy.” *International Political Science Review* 21:3, 261–283.

- Munck, Gerardo and Richard Snyder. 2007. *Passion, Craft, and Method in Comparative Politics*. Baltimore: Johns Hopkins University Press.
- Mukherjee, Siddhartha, 2010. *The Emperor of All Maladies*. New York: Scribner.
- Ragin, Charles. 2000. *Fuzzy Set Social Science*. Chicago: University of Chicago Press.
- Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64:4, 1033–1053.
- Sartori, Giovanni. 1984. *Social Science Concepts: A Systematic Analysis*. Thousand Oaks, CA: Sage.
- Sartori, Giovanni. 1993. "Totalitarianism, Model Mania, and Learning from Error." *Journal of Theoretical Politics* 5:1, 5–22.
- Solow, Robert. 1997. "How Did Economics Get that Way and What Way Did It Get?" *Daedalus* 134:4, 87–100.
- Van Evera, Steven. 1997. *A Guide to Methods for Students of Political Science*. Ithaca: Cornell University Press.
- Zerubavel, Eviator. 1996. "Lumping and Splitting: Notes on Social Classification." *Sociological Forum* 11:3, 421–433.

Against Methodological Unity: A Critique of Gerring

Patrick Thaddeus Jackson
American University
ptjack@american.edu

There is a paragraph in John Gerring's book with which I whole-heartedly agree. It is the last paragraph in the book proper, the conclusion of the Postscript:

We ought to begin with a recognition that social science constitutes an independent—although never entirely autonomous—realm of endeavor. The trick is to make social science speak to problems that we care about without sacrificing the rigor that qualifies it as a science. This is not an easy trick, but it is the trick of the trade. (p. 401)

The distinctiveness of the social-scientific endeavor strikes me, as it apparently does Gerring, as the appropriate starting-point for any attempt to figure out what social science is and what it should be. In particular, social science needs to be distinguished from social activism or social policy, lest what we academics do become nothing but another forum within which the same political struggles present elsewhere in society can be fought out, perhaps with different weapons, but with the same basic urgency and goal: victory, for one's preferred partisan perspective, whatever the cost.

I call attention to this paragraph because it is one of the only places in the book where I am entirely in agreement with Gerring's position. Virtually every other point he makes is, in my view, distinctly and perhaps fatally undermined by his steadfast refusal to brook any debate or compromise when it comes to his stipulated definition of "science": "systematic, rigorous, evidence-based, falsifiable, replicable, generalizable, non-subjective, transparent, skeptical, rational, frequently causal, and cumulative" (p. 2). The latter terms of this series cash out the former in ways that threaten to erode even my whole-hearted assent to that last paragraph about the distinctive-

ness of the social-scientific enterprise, because if "the rigor that qualifies it as a science" must be "falsifiable, replicable, generalizable, nonsubjective, transparent, skeptical, rational, frequently causal, and cumulative," then I can't in good conscience agree. Fortunately for me, and others like me, science is a heck of a lot broader and more diverse than Gerring makes it out to be, and if Gerring were truly interested in carving out a distinctive space in society for social science, he might want to consider not alienating his potential allies in pursuit of a dream of cumulative progress that most philosophers of science abandoned years ago.

Indeed, there is a curious out-of-time quality to Gerring's book, as though a portal to the 1950s had suddenly opened and it was possible to talk in pretty simplistic terms about falsifiability and nomothetic generalization as important cornerstones of a unified approach to social science. One finds this curious lack of engagement with important debates in the philosophy of science in other works by political scientists aiming at the social science in general, most notably in King, Keohane, and Verba's eponymous handbook—a lack stressed in his Section's 2006 article of the year (Johnson 2006), an article that is curiously not referenced by Gerring at all. Instead Gerring, like King, Keohane, and Verba, posits a view of science that valorizes hypothesis-testing as a way of producing approximately general laws along with a correlational view of causation, and then uses that as a position from which to critique other views of knowledge-production for not living up to those ideals. This is hardly likely to be compelling to anyone who does not already share that view of science, and given that the view on offer is *neopositivist*—lacking the philosophical subtlety of logical positivism, it treats as firmly established the very principles that the Vienna Circle logical positivists spent copious amounts of time wrestling with—the resulting argument is not going to win any support from critical realists, analyticists, or reflexive theorists. All of these have their own ways of being "rigorous" and "systematic" and "evidence-based," but all of them would (properly!) blanch at notions like replicability and generalizability—and in doing so they would have large numbers of contemporary philosophers of science on their side.

Philosophically speaking, the grounds on which Gerring wants to evict such scholarship from the cathedral of science is extremely shaky. I was really hoping for an actual *argument* about the need for a single, uniform methodology of social science; instead, I found the hoary old chestnut of "relativism," the straw-man construction of Kuhnian "incommensurability" as producing hermetically sealed research communities that can't even talk to one another, and the specter of craniology/phrenology as the kind of thing we need a firm demarcation criterion to exclude from the sphere of scientific knowledge. I expected more. Numerous studies in the (post-Mertonian) sociology of science (e.g., Latour and Woolgar 1986; Pickering 1995; Taylor 1996) have basically demolished the notion that any strict demarcation criterion for the boundaries of science exists or has any influence over the kinds of scientific knowledges that are accepted or rejected. Kuhn *himself* developed his notion of "incommensurability" beyond the ini-

tial statement in *The Structure of Scientific Revolutions*, ending up with a view that stressed the context-dependence of specific conceptual terms rather than the whole-sale incompatibility of broad schools of scientific research (Kuhn 2000). And as for “relativism,” what seems to be at issue in such charges is the fear that two different approaches to a topic or question will produce contradictory findings, but this *simply doesn't happen* because scientific approaches of all flavors and varieties are about producing what Dewey (Dewey 1938; see also Hickman, Neubert, and Reich 2009) called “warranted assertability”: claims that are justified by appropriate evidence. Such claims are either (1) translatable into another lexicon in which they will be similarly warranted albeit using different terms (as Nicholas Rescher [1997: 61] sardonically remarked, two shekels plus two shekels is four shekels even if you're not a Babylonian), or (2) literally incomprehensible. All sorts of intriguing problems of translation between different sets of warranted assertions crop up at this point (some of which are explored in Davidson 1973), but precisely *none* of them have anything to do with one scientific approach (whether we are talking about an academic discipline, a theoretical school, or whatever) losing its grasp on the world because another approach comes along and contradicts it.

Instead of a compelling argument, we simply get assertions that science has to be about consensus if we social scientists are to have anything useful to offer to the world. This leads Gerring in the direction of the kind of faux tolerance of the sort that one sees in David Laitin's kind of pluralism-that-isn't-pluralist (e.g., Laitin 2003; critically discussed in Jackson 2006): as long as everyone is on the same basic page, methodologically speaking, we can all get along. And as in Laitin, this drive to include leads to a (likely unintentional) misrepresentation of positions that—inconceivable though this might be for a committed neopositivist—actually have *different epistemic goals*, and aren't just trying to create nomothetic generalizations about cross-case covariation. Thus Gerring misreads causal mechanisms as intervening variables, even to the point of claiming that “the existence of a causal mechanism *presumes* a pattern of association between an exogenous X and an endogenous Y” and that a causal mechanism “presupposes a pattern of association among a set of intermediate variables” (p. 375)—all of which would likely come as a shock to non-neopositivist social scientists for whom the very notion of a causal mechanism is that it represents a dispositional property and is not therefore reducible to any finite set of observed associations. He likewise equates “qualitative” with “small-n” (p. 362) in ways that make the very notion of “qualitative” quite vacuous (if “qualitative” simply meant “small-n,” why would anyone use the term “qualitative” in the first place?), and extends “hermeneutics” the same backhanded compliment that King, Keohane, and Verba did a decade and a half ago: thinking about what people mean might help you generate some really interesting hypotheses, but hermeneutics can't possibly be an entirely different *methodology* (pp. 47–48). I am put in mind of a remark made by Immanuel Kant:

Differences in religion: an odd expression! Just as if one spoke of different *moralties*. No doubt there can be dif-

ferent kinds of historical *faiths*, though these do not pertain to religion, but only to the history of the means used to promote it... There is only a single religion, valid for all men in all times. Those [means used to promote it] can thus be nothing more than the accidental vehicles of religion and can only thereby be different in different times and places. (Kant 1983: 125)

Running through both Kant and Gerring is, I think, an unwarranted *fear of diversity*, as though any acknowledgement of the fact that science is plural and pluralist would bring the entire edifice of modern civilization crashing down around our ears. Gerring denies, by simply not acknowledging, that there are many different epistemic purposes pursued by many different sciences both social and natural, that there is no single set of philosophically defensible principles that governs a unified science, and that according to many contemporary philosophers of science—including Larry Laudan, cited numerous times in the text but not on this point—*this is perfectly okay*. Gerring never takes on Paul Feyerabend's argument that science only discovers anything new by having a plurality of claims against which new speculations can be compared. Instead, he continually alludes to “us,” “our” everyday judgments, “our” common intuition about what makes a good argument, without ever really specifying who “us” is—and without justifying the metaphysical leap from “what we all think” to “what is in some broader sense warranted or justified.” (Kant, to his credit, was incessantly worried about just this leap.) Gerring never makes what Colin Elman (2009) aptly calls the “hard choice” for an engaged pluralism (see also Lapid 2003). Instead, Gerring's goal is to find consensus, even to the point of denying that fundamental epistemic diversity exists.

The great irony is that Gerring's narrow methodological tent actually harms what might be seen as his broader project of holding open that space for the social sciences to pursue their own distinctive vocation. Gerring's ultimate claim about the value of scientific consensus seems to be that we have nothing to offer the public if we don't have a unified methodology that can generate claims which command general consensus among the community of scholars (p. 10). This is a somewhat naïve, and indeed, very *American*, understanding of the proper relationship between science and politics: politics is messy, science withdraws from politics so as to gain some measure of truth that it can then use to correct and reform politics (Gunnell 1993). But scientific consensus doesn't seem particularly politically decisive these days, even and perhaps especially in the United States; one need only think of current political debates about global climate change that proceed *despite* the overwhelming scientific consensus on the anthropogenic aspects of such changes, or the complete lack of impact that the overwhelming consensus of the professional International Relations scholarly community had on the course of U.S. military policy in Iraq (Jackson and Kaufman 2007). On the other side of the coin, the popular use of the “democratic peace” as a justification for all kinds of democracy-promotion activities presumes a scholarly consensus that in fact doesn't exist, and also badly distorts the argument for political purposes. So it would seem that what social scientists have to

offer the public is highly contingent on what the public is interested in, which has little or nothing to do with actual scientific consensus.

In fact, there are other views of the relationship between science and politics that, arguably, capture that relationship better than the optimism of the progressive reformer: I am thinking primarily of Max Weber's admonition that social science's vocation is to clarify the likely implications of decisions but *not* to take over the responsibility for those decisions (Weber 2004). Weber's call for scientific value-clarification rather than scientific value-correction does not demand a broad-based scientific consensus, and seems better suited to a world in which we *already* have a diversity of ways of doing science. But such methodological Weberians, by virtue of not consenting to the value of scientific consensus, are excluded from being able to help advance Gerring's goal of a conditionally autonomous social science.

Further, embracing the diversity of science would also help another of Gerring's declared goals: the pragmatic goal of helping us see what positive contributions social science can make.

Rather than choosing camps, let us ask what specific tasks, strategies, and criteria each camp entails. We can then ask the question: would the social sciences, thus oriented, tell us about things that we want to know? (p. 398)

Unfortunately, Gerring then goes on to ask about the promotion of "societal consensus," seemingly unaware that by limiting his criteria to criteria of consensus he is precluding us—and by "us" I mean the community of social scientists, broadly defined—from asking questions about the value of consensus itself. But why limit the discussion arbitrarily? Why not let social science, broadly defined as systematic inquiry into the dynamics of the social world, unfold in myriad of ways? If our value as social scientists is, in the end, that we produce knowledge that is in some sense valid and therefore useful, I would think it extremely short-sighted to place any *a priori* limits on that process. Pluralism helps us all, by making sure that all the possible avenues are investigated, and I for one want to see how things turn out rather than pretending that I know the answers in advance.

References

- Davidson, Donald. 1973. "On the Very Idea of a Conceptual Scheme." *Proceedings and Addresses of the American Philosophical Association* 47, 5–20.
- Dewey, John. 1938. *Logic: The Theory of Inquiry*. New York: Henry Holt and Company, Inc.
- Elman, Colin. 2009. "Letter from the Section President: Pluralism as a Hard Choice." *Qualitative and Multi-Method Research* 7:2, 1–2.
- Gunnell, John G. 1993. *The Descent of Political Theory: The Genealogy of an American Vocation*. Chicago: University of Chicago Press.
- Hickman, Larry A., Stefan Neubert, and Kersten Reich, eds. 2009. *John Dewey Between Pragmatism and Constructivism* (American Philosophy). 3rd ed. New York: Fordham University Press.
- Jackson, Patrick Thaddeus. 2006. "A Statistician Strikes Out: In Defense of Genuine Methodological Diversity." In *Making Politi-*

- cal Science Matter: Debating Knowledge, Research, and Method*. Sanford Schram and Brian Caterino, ed. (New York: New York University Press), 86–97.
- Jackson, Patrick Thaddeus, and Stuart J. Kaufman. 2007. "Security Scholars for a Sensible Foreign Policy: A Study in Weberian Activism." *Perspectives on Politics* 5:1, 95–103.
- Johnson, James. 2006. "Consequences of Positivism: A Pragmatist Assessment." *Comparative Political Studies* 39:2, 224–252.
- Kant, Immanuel. 1983. *Perpetual Peace and Other Essays*. Indianapolis: Hackett Press.
- Laitin, David. 2003. "The Perestroika Challenge to Social Science." *Politics and Society* 31:1, 163–184.
- Lapid, Yosef. 2003. "Through Dialogue to Engaged Pluralism: The Unfinished Business of the Third Debate." *International Studies Review* 5:1, 128–131.
- Latour, Bruno and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton: Princeton University Press.
- Pickering, Andrew. 1995. *The Mangle of Practice: Time, Agency, and Science*. Chicago: University of Chicago Press.
- Rescher, Nicholas. 1997. *Objectivity: The Obligations of Impersonal Reason*. Notre Dame, IN: University Of Notre Dame Press.
- Taylor, Charles Alan. 1996. *Defining Science: A Rhetoric of Demarcation*. Madison: University of Wisconsin Press.
- Weber, Max. 2004. *The Vocation Lectures*. Indianapolis: Hackett Press.

Response: Fear and Loathing on the Methodological Trail

John Gerring
Boston University
jgerring@bu.edu

Methodological textbooks—at least those that try to do more than impart statistical techniques—come in two genres. The first genre is all-embracing, including all manner of work conducted in a discipline. These sorts of texts elaborate diverse traditions of inquiry but they do not tell the reader very much about how to prioritize different approaches and do not elucidate common ground that might unite various methods. This sort of a textbook may be better classified as sociology or history rather than as methodology, for it does not elucidate good/bad practices (or does so only to a very limited extent).

The second genre of methodological textbook propounds a unified framework and is fairly specific about what constitutes good/bad practices. Naturally, it is a bitter pill for those who disagree with it. But it is in keeping with the prescriptive goal of methodology. It endeavors to define what good/bad social science is and, along the way, imparts practical "dos and don'ts."

My own view is that SSM2 could be classified in either genre. Sometimes, the book reads to me like a mealy-mouthed "everything goes" textbook that has something for everyone and is anxious to please all customers. At other times, it seems overbearing and opinionated.

The participants in this symposium seem largely of the latter opinion, which is why I have titled my contribution "Fear and Loathing" rather than "Goldilocks and the Three Bears."

(The notable exception is Dafoe, who admonishes me to provide more explicit arguments about what to do and what not to do.) My other reason for choosing this title is the brevity and crudeness with which I shall deal with the many subtle points raised by the contributors, more than a little reminiscent of Hunter Thompson's "Gonzo" journalism.

In any case, I hope that the tenor of the book, and of my discussion in these pages, is forthright and respectful. Sometimes, progress is achieved by making a clear argument, even if that argument is perceived as exclusive rather than inclusive.

Before launching into these subjects I want to thank the participants for contributing to this symposium and for articulating their views at some length and with great eloquence. It is an honor for an author's work to be the subject of probing commentary, and I look forward to further exchanges in the future. I suspect that all I shall have time for in this venue is setting down a few markers for future debate.

These markers will focus largely on the critical comments in the symposium, leaving aside points of agreement or orthogonal comments. Since there is some overlap in the contributions I will organize my thoughts according to topic rather than author (even though each section relates primarily to a single author). The topics: (a) complexity, (b) concepts, (c) the experimental template, (d) natural experiments, (e) mechanisms and causal effects, (f) set-relational and potential-outcomes causal models, (g) monism and pluralism, (h) what is social science?, (i) a best method?, and (j) other issues?

Complexity

Derek Beach correctly notes that the new edition of SSM is a lot longer and more complex than the previous edition. (Yes, longer AND more turgid! All for one low price.) Whether it is a good point of departure for beginners is a question that instructors will need to consider.

I hope to have written a book that does not require much prior knowledge. But clearly, it requires a good deal of engagement with methodological topics and that is something that not all graduate students—needless to say, not all undergraduates—can muster. (I was one of those students.)

That said, I feel strongly that the topics undertaken in this book are ones that every graduate student should be exposed to. The criterion for inclusion in SSM2 might be summarized as "things I wish I had learned in grad school (but did not due to my thick-headedness)." My sense is that most of the topics broached in SSM2 have now become part of the required skill-set of the discipline.

Allan Dafoe suggests that the book could be made more tractable if the many criteria laid out in the course of the book (and summarized in Table 1.1) were reducible to a smaller number of categories—categories that could be considered basic to the scientific enterprise. The book would then be more parsimonious and also more coherent (centrally important criteria, according to SSM2).

Two *Ur*-categories are suggested in SSM2 (chapter 2): (a) discovery and (b) appraisal. These are, of course, well-worn categories in philosophy of science, harking back to Reichenbach (1938), and perhaps further. They also exemplify an en-

during tension in the work of science, exemplified in the (rather extreme) positions adopted by Feyerabend and Popper, respectively. More prosaically, there is also an old ditty that I rather like (and that my students find amusing): "If true, not new; and if new, not true." That sums up a lot of methodological debate. In any case, these two goals motivate many of the more specific criteria discussed in SSM2 and are therefore foundational. (Perhaps I could have done a better job of illustrating these interconnections.)

Dafoe nominates (a) testability, (b) consistency, and (c) usefulness as *Ur*-categories. The first is pretty much equivalent to appraisal, but the others are a little different. I don't think I would view consistency as foundational, though it is certainly important (see chapter 3). Usefulness is a no-brainer, but then again it's hard to evaluate because it has so many possible meanings—theoretical utility, conceptual utility, practical/policy utility, everyday utility (telling people things they wish to know about). (Often, these desiderata are in conflict with each other.) Indeed, the entire framework of SSM2 rests on a pragmatic foundation, as explained in the Postscript. I argue that the sort of social science we should cultivate is the sort of social science that will help us answer questions that we wish to understand and address problems that we wish to solve. This is usefulness in the broadest sense and it is critically important—in my view—to any argument about what social science is, or should be (more on that below). So, I embrace usefulness wholeheartedly but with the caveat that in its broadest sense it applies to everything and in its narrower sense it only applies to some things. A final issue to consider is whether Dafoe's three categories encompass all (or even most) of the criteria in SSM2. I think it might, but only if usefulness provides a catch-all residual category.

In any case, thinking about how the criteria fit together—and how they might be reduced—is a worthwhile endeavor and one that I hope to think through again if and when I revisit this text. (As Allan alludes, I am currently considering the possibility of writing a shorter, simpler text that would be designed primarily for undergraduates.)

Concepts

It is fair to say that SSM is a melding of different traditions of work on concept formation, including both the more deductive, logical tradition of Giovanni Sartori (continuing a tradition founded by Frege and the logicians) and the empirical tradition stemming from work on measurement. I appreciate Ariel Ahram's comments on the applicability of part II of the book, focused on the—oft-neglected—topic of Description.

The Experimental Template

King, Keohane, and Verba (hereafter "KKV" 1994) is often accused of adopting a "quantitative template" for the work of social science. If SSM2 attains even a smidgen of the notoriety that KKV has attained it may also come to be summarized in a single moniker. And, if so, that moniker will probably be the "experimental template." Of course, no author likes their work to be reduced to a slogan, but if a slogan must be found this is probably as good as any other. (It is a somewhat ironic slogan,

as I have to date not conducted a single experiment.)

The question, then, is how appropriate—how useful—is this template? Beach has doubts. In particular, he has doubts about whether this view of research design applies to case study research. He writes: “most students will be left with the impression that big [in the sense of a larger N] is (almost) always better...and that we use case studies only when we are *unable* to do large-n research (preferably in the form of experiments).”

SSM2 treats case study methods fleetingly, partly because I have a separate text on that subject (Gerring 2007a). But the framework developed in SSM2 is intended to encompass case study methods, so a few words on the latter may be justified (see also the discussion of causal-process observations in chapter 12). Briefly, it is true that case studies may be justified as the best possible research designs when large-N quasi-experimental designs are impossible or improbable (i.e., the assumptions necessary to reach causal inferences using such designs are not plausible). It is also true that case studies are often useful as complements to a large-N analysis, a mixed-method approach.

Are experiments the gold standard? I think this truism is true—more or less, with a few caveats, as reviewed in SSM2 (e.g., problems of confounding that creep in after the assignment of a treatment, problems of causal heterogeneity, and problems of external validity). Moreover, I believe that many critics of KKV accept this. Certainly, this is a guiding principle of *Rethinking Social Inquiry* (RSI) (Brady and Collier 2010), of David Freedman’s oeuvre (2001, 2010), and—I think—of most others cited by Beach in this passage. Indeed, this line of critique of KKV is not about having adopted an experimental template but rather about not remaining true to the experimental template. Jay Seawright’s chapter in RSI (2010) makes this point abundantly clear. The problem is not experiments; the problem is “regression”—a shorthand expression for large-N studies based on observational data. Sometimes, KKV is understood as a license to practice regression-based causal inference whenever a large sample can be enlisted. I don’t know if it is a fair reading of the book but that seems to be what many folks see as the “take away.” It is this model of non-experimental large-N analysis that is under attack—by statisticians and by case study researchers. In my view, this attack has perhaps gone a bit too far in the direction of methodological purity (see Gerring 2011). But I agree with the broad outlines of the critique. It is the experimental or quasi-experimental quality of the data-generating process, not the number of data points, that determines the ease and certainty with which causal inference can be achieved. Note that this is a primary selling-point for case study research: Sometimes, by looking at one or two cases, one can more readily avoid confounders than when examining hundreds or thousands (Gerring 2007a; Gerring and McDermott 2007.)

Beach wonders, however, whether the experimental standard is relevant to case study research whose objective is not the generation and testing of a general theory but rather the explanation of a single outcome (“causes of effects”). He writes, “If we are interested in studying why India is democratic de-

spite everything we know from other cases about the relationship between economic development and democracy, is the first thing we ask ourselves ‘How can I test a hypothesis that could explain the deviance in an experimental fashion?’”

Well, first you need a hypothesis, and no experiment can be conducted until you have one. In this sense, observational analysis (large- or small-N) and head-scratching always come first. But once you have a hypothesis, I would say that your first recourse should be to try to devise an experimental or quasi-experimental design. If the latter is impossible—as it often is in historical research and research with macro-level institutions that are not directly manipulable—then one must fall back on other expedients. This seems self-evident to me. All I am saying is that we have an obligation to use the best tools available. And where the best tools are not available we need to understand the limitations of the tools that we use. More on this below.

Natural Experiments

Dafoe is unsatisfied with SSM2 insofar as the text does not clearly lay out a way forward, i.e., a better way of doing social science. (Unlike the other participants in this symposium, Dafoe thinks SSM2 is not prescriptive enough.) In particular, he wishes I had more to say about design-based inference—which I take to be a synonym for *natural experiments*. I am glad to address this question because it’s both important and complex.

First off, let me say that I wholeheartedly embrace the current methodological trend—emanating from work by Donald Campbell and others (e.g., Rubin 2008)—to emphasize *ex ante* design over *ex post* analysis. See SSM2 (pp. 78–80) for a strong statement of this position. I am also a fan of Thad Dunning’s work. (Thad’s book, *Natural Experiments in the Social Sciences: A Design-Based Approach*, will appear in the same Cambridge UP series—*Strategies for Social Inquiry*—later this year.)

However, I shied away from the terminology of natural experiments in SSM2 because I do not view natural experiments as fundamentally distinct from other observational studies. The usual definition of a natural experiment is that assignment to treatment is at random or as-if random. This condition may be achieved through a simple comparison of cross-group means (no adjustment necessary), by regression-discontinuity (RD) designs, or by instrumental-variable (IV) designs. However, the claim of *as-if randomization* is a claim about an actual data-generating process that cannot usually be verified, at least not in any strong sense. Thus, unlike an experiment, the assumptions necessary to causal inference are quite important and quite ineffable. They may be true, and they may not be. Indeed, some invocations of the term strike Dunning and other observers as rather farfetched. “Natural experiment” presents a terminological boundary that is difficult to police. This seems especially true of the incessant use of IV designs. One person’s natural experiment is another’s “messy observational study.”

What this suggests to me is not that the ideal is flawed; indeed, it is the experimental ideal that all causal inference

rightly strives for. It suggests, rather, that the presumed distinction between natural experiments and observational data analysis is not bright and clear. Likewise, the assumptions needed to achieve causal inference in some so-called natural experiments are less plausible than the assumptions needed to achieve causal inference in some observational studies that do not invoke this term (i.e., where lots of covariate-control is applied). It is the nature of these assumptions—and the convincing-ness of various robustness tests—rather than a study's status as a natural experiment that proves decisive. If so, the term may obscure as much as it clarifies.

Another reason for shying away from the term is that it emphasizes only one—albeit incredibly important—aspect of research design: assignment to treatment. Insofar as research design involves lots of desiderata (discussed in painstaking detail in chapters 4, 9, 10, and 11), it was important not to adopt a term that focuses only on one aspect of the problem. (I have the same misgivings about “experiments,” although in this case the crispness of the category suffices to justify its use.)

To conclude: I am a proponent of design-based inference. But I think we need to think about the design properties of a study along more than one dimension. With respect to natural experiments, if they are narrowly and strictly interpreted they are likely to remain a highly restricted tool in the arsenal of social science (much more restricted than true experiments). There just aren't enough serendipitous situations out there to capitalize on. And when they are found, they are often hard to generalize from. If the term is loosely interpreted, natural experiments are more ubiquitous; however, this species of natural experiment should not be privileged above other observational studies. And since the strict/loose interpretation is not easy to specify—since it depends upon largely untestable assumptions—I decided to downplay the term in SSM2.

Mechanisms and Causal Effects

SSM2 emphasizes the search to identify causal effects as well as the search for causal mechanisms. (More ink is spent on the former in SSM2 simply because the methodology is a lot more advanced, a feature of the social science world that may be changing.) I do not see this as problematic. Indeed, I do not see it as avoidable. As discussed in chapter 13, there are few causal theories in the universe of social science that are purely “covariational” or purely “mechanismic.” Most causal theories combine these objectives. They seek to demonstrate what impact on *Y* a change in *X* has (or had) and they seek to explain why *X* has that impact on *Y*. Insofar as this is an enduring tension in social science research it is a productive tension—because we want both sorts of knowledge.

Nor do I think that the experimental template is at variance with the desire to elucidate causal mechanisms (as Beach implies). Indeed, many experiments are devised for the purpose of testing the mechanisms operative in a well-established causal relationship (Glynn and Quinn 2011; Imai et al. In press).

Beach argues that mechanisms are not amenable to a manipulationist view of causation because they are continuous processes, a holistic system. This is a big debate among philosophers of science (Nancy Cartwright's work in support

of this view should be added to those that Beach cites), and not something I can deal with adequately here. Nor is it dealt with in SSM2 (I tried to pass over philosophical debates as much as possible in order to retain the book's appeal as a teaching tool).

Briefly, the case for a manipulationist view of mechanisms goes something like this. First, a mechanism is an arbitrary construction—anything that lies between “*X*” and “*Y*.” Change the *X* or *Y* and you change the mechanism. The accordion can stretch out (in which case there are more mechanisms lying between *X* and *Y*) or it can scrunch together (in which case some of the phenomena are defined out of the model). However *X* and *Y* are defined, the mechanism—*M*—is the stuff that is left in between. It might be a micro-mechanism, e.g., a chemical process. It might be a macro-mechanism, e.g., a process occurring at the level of national or international institutions. The time period might be brief (a nanosecond) or long (millennia). Now, if mechanisms are truly special—vis-à-vis *X* and *Y*—how can it be that their special-ness is so arbitrary? If a phenomenon is manipulable as a cause (*X*), why is it not manipulable as a mechanism (*M*)?

Second, to specify a mechanism is to make a causal claim with an implicit counterfactual. That is, to say that *M* is a mechanism lying between *X* and *Y* is to say that if *M* is removed or changed in some fashion, the relationship between *X* and *Y* will be altered. The counterfactual is *M*'s removal or alteration. This counterfactual implies a possible intervention, which may be manipulated in some fashion. Of course, we are not saying that all mechanisms are amenable to experimental manipulations. But neither are all causal factors (*X*). All we are saying is that, in principle, it ought to be possible to devise a manipulation that will alter the condition of a mechanism such that it no longer serves its mechanismic function.

For these reasons, I think it is possible to bring mechanisms into the world of causal inference as understood through the potential outcomes framework. (For further work on this trajectory see Glynn and Quinn 2011; Imai et al. In press; Pearl 2009).

Of course, simply measuring and conditioning on a mechanism is not sufficient to tell a causal story (Gerring 2012). Something more is required, in particular, assumptions about the way the world works, and perhaps an explicit theory about how *X* causes *Y*. No amount of empirical testing will obviate this fact, and Beach is correct to mention it. However, this is a separate issue from manipulability. I think the confusion arises because the term “mechanism” has come to carry so much semantic freight (Gerring 2007b). In particular, it sometimes refers to intervening variables and sometimes to the causal story or theory that one tells about *X*'s relationship to *Y*. We should keep these things separate, as much as possible. There is nothing implicit or explicit in the manipulationist account that should lead one to downplay the role of theory or of causal assumptions about how the world works.

Set-Relational and Potential-Outcomes Causal Models

This is the most detailed contribution and it really requires a lengthy response, lengthier than I have time or space to

provide. I think the question of how set-relational (SR) and potential-outcomes (PO) models of causation relate to one another is absolutely critical, and pretty much ignored in the literature. If a unified framework of causality is to succeed, it must encompass these two traditions.

Before beginning, let me take note of my own ambivalence with respect to the potential outcomes model. As discussed in chapter 12, the PO model now means many things to many people and so it has—like SR—become a difficult matter to generalize about. I think that an encompassing view of PO has the potential to apply broadly to work in the social sciences; in that sense, I am optimistic about the project sketched by Morgan and Winship (2007) and many others. However, like most tools, it also has limits, which is why I scarcely mention the term in SSM2 and don't see myself as committed to it. That said, let me approach the Schneider/Rohlfing (hereafter S/R) challenge as a challenge worth thinking about, and one that is suggested briefly in my treatment of set-theoretic work in chapter 12. (I am grateful to Carsten for his assistance in writing that portion of the chapter.)

S/R begin by relating my argument about how the PO and SR approaches can be understood as complements, rather than antagonists. My argument, in a nutshell, is that they are attempting to understand *different elements* of a causal relationship. PO is focused on the causal effect—the impact of a change in X on Y relative to what Y would otherwise be (the counterfactual condition). SR is focused on the probability of a set of outcomes for Y given a value for X (where X may include a vector of factors scored in a binary fashion). Specifically, can this probability be understood as necessary or sufficient?

Before going any further, let me clarify that SR relationships may be important even when they are not causal. If I know that whenever $X=1$, $Y=1$ (or whenever $X=0$, $Y=0$), this could be very useful information, even if the relationship is not causal. Here, there is no intersection between PO and SR because PO is focused exclusively on causal relationships. So, the following discussion presumes that we are talking about the use of SR techniques in situations where the researcher is interested in necessary or sufficient *causal* relationships.

A small point (well, perhaps not so small): I don't see myself saying that “the p-score of the ATE is the *only* basis for inferring causality.” If I do imply this somewhere, I certainly don't mean to. A key message of SSM2 is that causal inference is not inferable from statistics. It is inferable from what we know about the world and, relatedly, from our judgments about the soundness of whatever assumptions are needed in order to believe that the causal model mirrors (in relevant respects) the actual data-generating process.

What I say, or at least what I mean to say, is that the investigation into a causal effect (understood through the PO model) is a useful supplement to an investigation into SR relationships. I presented it in the book as a precursor. But, really, it could happen at a secondary stage. It doesn't much matter. The point is that if X is necessary or sufficient for Y it should be possible to locate a causal effect and, depending upon how convincing that evidence is, it will bolster our willingness to believe that X is necessary or sufficient for Y . If, on the other

hand, no such causal effect can be found, which is to say that the pattern of observed data is not distinguishable (at a fairly high level of certainty) from what might have happened if X were not a cause of Y , then we are much less inclined to believe any arguments that might be made about necessity and sufficiency.

Let us begin with an experimental setting, where the points are easiest to demonstrate. Suppose, following S/R, that our treatment and outcomes of interest are binary. And suppose that $X=1$ is (causally) sufficient for $Y=1$ in some population of units where there is variation in Y . It is axiomatic that if the treatment ($X=1$) is randomly assigned across two groups, the group receiving the treatment will have a higher number of $Y=1$ outcomes than the control group. (In fact, all the outcomes in the treatment group will be $Y=1$, but this is beside the point for present purposes.) Likewise for any necessary relationships, for all units in the control condition ($X=0$), $Y=0$, while units in the treatment condition will exhibit a mixture of outcomes. In both cases, a treatment effect should be easy to demonstrate—assuming a reasonably large sample.

With observational data, things are more complicated (as always). But it is still the case that one's willingness to believe that a necessary relationship is causal is bolstered if one can discover a treatment effect in the data, and mitigated if one cannot. (I leave aside software issues that arise when probabilistic statistical protocols are faced with exception-less SR relationships.) So, it is true that, as S/R write, “the ATE can be insignificant in the presence of a sufficient condition—a *false negative*—and...the ATE can be significant in the presence of a non-sufficient condition—a *false positive*.” I did not mean to imply that finding a causal effect was a precondition (necessary condition) of finding a necessary or sufficient causal relationship. Rather, the point was that finding a causal effect would bolster our faith in the SR claim. In particular, it would bolster our faith in any SR argument that has a fairly high level of “coverage”—which, one might argue, are precisely the sort of SR arguments that we ought to be worrying about. The *degree* to which an investigation of causal effects updates our priors about an SR relationship is of course dependent upon the strength of the research design, its external validity, and all the things that we would normally be concerned about.

In Table 4 and accompanying text the point S/R make is that a causal effect does not discriminate between necessity and sufficiency. True. But that is not its purpose. Its purpose is to help demonstrate that a relationship is causal, rather than merely correlative.

In Table 5, the point is that if a sample is too small a probabilistic test such as a t test will incorrectly reject a true causal relationship. Or it may correctly reject a false relationship. From my perspective, if one is basing one's conclusions about causality on twelve cases one must be fairly tentative about what the relationship actually is. Again, the point of an investigation into causal effects is to evaluate the hypothesis against a counter-hypothesis, the so-called null hypothesis. This hypothesis might be something different; the PO model is not wedded to frequentist models of inference. But there is a possibility of failure and it is—hopefully—one that can be evalu-

ated against a reasonable model of how the world works.

At this point, the question before us becomes the following: Do tests of consistency and coverage provide tests of validity, sufficient to displace tests based on the PO model? I think not. But I am hesitant about this judgment because I am not clear on the body of knowledge that would support the assertion that, for example, if a relationship surpasses a given level of consistency and coverage, it is (causally) sufficient or necessary. Note that models used to evaluate causal effects have a clear counterfactual, either a null hypothesis or some other. Thus, surpassing that threshold—provided one accepts the assumptions of the model—has a clear meaning. I am not clear on what meaning to ascribe to tests of consistency and coverage.

So, to conclude, my position is that work on SR suggests that in order to determine whether there are necessary or sufficient causal relationships one should look at the distribution of cases across the matrix (which might be quite a bit more complicated if the relationships are interactive, as QCA relationships are). I am suggesting that, in addition, one should look for causal effects. If these effects confirm the discovered SR relationships, one is more inclined to believe that the latter are actually causal (contingent upon the strength of the research design used to assess causal effects).

I realize, of course, that this is not always possible. Sometimes, experiments are not possible and available observational samples may be limited in size. Sometimes, the sheer number of possible combinations overwhelms the degrees of freedom offered in a limited sample. But, in these circumstances, circumspection is warranted. In other words, in precisely those instances where investigating a causal effect is difficult, one should be cautious about accepting an SR inference as causal simply on the basis of the covariational distribution of cases.

Now let me add an essential point that I know S/R agree with. Causal inference in SR work rests largely on in-depth case study analysis. It is unclear whether, or to what extent, case study analysis can be fruitfully explored with the PO model. And it is undeniable that case study evidence rarely allows for the identification of causal effects. So, in these respects the PO/causal effect approach to SR will not help. If we think of SR work as resting upon three legs, one leg is the distribution of data across the matrix, as illustrated by S/R and as assisted (in complex large-N settings) by the Boolean QCA machinery developed by Charles Ragin. Another leg is case study research into causal processes operating through time (process tracing). And a third leg, I have tried to argue, is the investigation of causal effects.

A few more points probably deserve attention such as the distinction between causes of effects (traditional for SR) and effects of causes (traditional for PO). While these varying purposes may be said to distinguish most SR and PO investigations, we should not conclude from this that they constitute a hard dichotomy. Note that SR can be used, and is sometimes used, to elucidate effects of causes. Likewise, PO can be adapted to effects-of-causes investigations if the latter are understood as bundles of effects-of-causes investigations.

The larger point at issue here is that the world of PO and

SR are not alternate universes of causality, as they are sometimes perceived. In principle, they are easily reconciled because they are looking at different, but complementary, features of a causal relationship. In practice, they may sometimes assist one another in reaching causal inference.

Monism and Pluralism

The big argument of the book is about the unified nature of social science research. Of course, not all agree with this goal. Beach and Jackson, and probably Schneider and Rohlfing, are among them.

Let me briefly restate the arguments for unity. First, we need to be able to talk to one another across disciplines and subfields and methods. Without some sort of methodological agreement, we have no means of doing so and thus knowledge cumulation cannot occur. Second, from a purely pragmatic perspective, we need to make decisions about what is good research and not-so-good (or terrible) research. In doing so, we must appeal to norms that transcend “ontologies.” If ontologies (hard to define but let’s say “what we think is really going on out there but which is not empirically testable”) guide these debates then they are a matter to which no common ground can be found outside of the ontology. We are left making statements like “within my ontology, I am correct.” I don’t see this as very productive. Nor do I believe it is necessary. I believe there are norms that we all (or most of us) agree on, or could agree on: this is the framework laid out in SSM (summarized in Table 1.1).

This case is laid out in much more detail in chapter 1 and I shall not attempt to reproduce it here. Not everyone will be convinced. But I suspect that those who are not convinced will have trouble arguing for their point of view. Consider this statement of Beach’s: “adopting a unified logic results in a narrowing of the types of research questions we can legitimately ask, and masks the comparative advantages of non-experimental methods.” Well, what are the questions we can legitimately ask, and how would one assess “legitimacy” in this context? It is a bit like the argument for moral relativism. In order even to make that sort of argument one must assume a non-relativist (“objective”) position.

Likewise, if we adopt a pluralist perspective the question arises: How shall we decide what to publish, whom to hire and promote, and what to teach our students? Are there any boundaries at all to the (acceptable) practice of social science? There is a logical absurdity to pluralism if it is understood to encompass anything and everything that people who call themselves social scientists do. Most pluralists are not that pluralist. That is why I raised the hackneyed example of craniology/phrenology. But if most pluralists—presumably, including Jackson—wish to erect boundaries around what constitutes acceptable and unacceptable social science then they must have some principles that they appeal to in doing so. This is precisely the exercise SSM2 engages in.

Of course, there are degrees of pluralism, and I am not sure (though very curious) about where Jackson would define his own version. One can imagine a series of concentric circles with those things that are least acceptable (according to one’s

definition of social science) at the outer edge and those most acceptable at the center. Some people would draw the circle more tightly, others more broadly. But everyone, I submit, would draw a circle somewhere and would appeal to universal (trans-disciplinary) standards in doing so. From this perspective, SSM2 is no different from every other methodological textbook that purports to address normative concerns (how we should practice social science).

In short, while I am open to the charge of not having adequately framed the common and essential points of agreement across the social sciences, I find it rather unconvincing to argue that there are no common points of reference. And I should also note that some of the folks cited as exponents of “multiple-causation” don’t actually embrace the idea. It is one thing to point out that people have different ideas of causation and another to endorse pluralism as a way forward. Brady (2002) elucidates the four types of causation and then concludes by saying that they are not really so contradictory as they appear, suggesting that common ground is possible—and desirable. Mahoney’s most integrative study (2008) is titled “Toward a Unified Theory of Causality.” So, in my view, the fruitful debate is not whether we should have a unified framework but what sort of unified framework we should have.

What is Social Science?

In Jackson’s commentary a fundamental disagreement is articulated—one that might warrant the term “epistemological.” Jackson disagrees with the way in which I have defined science (and by extension social science), as “systematic, rigorous, evidence-based, falsifiable, replicable, generalizable, nonsubjective, transparent, skeptical, rational, frequently causal, and cumulative” (SSM2: 2). Evidently, if one doesn’t agree with this definition of science then not much in the book is likely to make sense. Indeed, it will be positively noxious. As I acknowledge later in this chapter, “my understanding of social science will not please everyone, and those unhappy with the point of departure are unlikely to be happy with the point of arrival.”

Jackson is correct to note that I stipulate this definition of science, rather than arguing for it. Granted, there is further discussion in the Postscript, but it is not the sort of extended philosophical argument that Jackson feels is required. The problem is that to offer a full defense of my preferred definition of science would involve a very different kind of book, a book of philosophy rather than a book of methodology. Moreover, the defense of scientific (in the sense adumbrated above) social science has been handled by philosophers (e.g., Blaug 1980; Hausman 1992; Kincaid 1996; Laudan 1996; Mayo 1996) more ably than I would have been able to do.

It is of course true that many philosophers of science have doubts about the view of science adopted in this book. But I would not take this as evidence that these arguments are true. There is a feature that one finds in a good deal of anti-positivist or post-positivist writing that surfaces in Jackson’s piece such that philosophical issues are considered to have been “settled by philosophers.” This is not my reading of the literature. Certainly, many well-respected philosophers of sci-

ence resist the post-positivist program. Anyway, I don’t want to pretend any sort of authority on this topic; I simply want to problematize some of Jackson’s—to me, rather blithe—assertions.

A Best Method?

Beach says that we need to advise our students that “(1) research questions determine the choice of method, not vice versa, and (2) they should choose the research method that offers the most bang for the buck in relation to a specific research question, i.e., it has the strongest comparative advantage in relation to offering analytical leverage in the particular research situation.”

I agree with (1) but with the caveat that the method of analysis is not entirely beside the point. Consider the well-worn parable about the drunk who looks repeatedly for his keys under the lamppost because that is the only area he can properly see. Sometimes, the discipline of political science seems to follow this logic, and it is troublesome. But it is not utterly devoid of sense. After all, if the drunks (political scientists) are finding a few keys under the lamppost every now and again, this is better than searching vainly in the dark—where there are, let’s say, lots of keys but we have an extremely low probability of discovering them.

I agree with (2) wholeheartedly. Indeed, it is the driving message of the concluding section of the last full chapter (chapter 14). The title of that section is “Best-Possible, All Things Considered.” I am sorry if this message was not clearly articulated prior to that point because it is the leitmotif of the book.

Other Issues?

Are there additional issues to debate with respect to SSM2? Of course! While this symposium has focused on a handful of important issues, I can imagine a whole category of additional issues arising from a different (more “positivist”) vantage point. For another day.

References

- Blaug, Mark. 1980. *The Methodology of Economics, Or How Economists Explain*. Cambridge: Cambridge University Press.
- Brady, Henry E. 2002. “Models of Causal Inference: Going Beyond the Neyman-Rubin-Holland Theory.” Unpublished manuscript. University of California, Berkeley.
- Brady, Henry E. and David Collier, eds. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd ed. Lanham, MD: Rowman & Littlefield.
- Freedman, David A. 1991. “Statistical Models and Shoe Leather.” *Sociological Methodology* 21, 291–313.
- Freedman, David A. 2010. *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. David Collier, Jasjeet Sekhon, and Philip B. Stark, eds. Cambridge: Cambridge University Press.
- Gerring, John. 2007a. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Gerring, John. 2007b. “The Mechanistic Worldview: Thinking Inside the Box.” *British Journal of Political Science* 38:1 (January), 161–179.
- Gerring, John. 2010. “Causal Mechanisms: Yes, But...” *Comparative Political Studies* 43:11 (November), 1499–1526.
- Gerring, John. 2011. “Large-N Observational Data Analysis (aka

- Messy Data): A Modest Defense,” *Qualitative and Multi-Method Research* (Spring 2011) 8–17, 31–33.
- Gerring, John. 2012. “Mere Description.” *British Journal of Political Science* (forthcoming).
- Gerring, John and Rose McDermott. 2007. “An Experimental Template for Case-Study Research.” *American Journal of Political Science* 51:3 (July), 688–701.
- Glynn, Adam N. and Kevin M. Quinn. 2011. “Why Process Matters for Causal Inference.” *Political Analysis* 19:3, 273–286.
- Hausman, Daniel M. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
- Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. In Press. “Experimental Designs for Identifying Causal Mechanisms.” *Journal of the Royal Statistical Society Series A*.
- Kincaid, Harold. 1996. *Philosophical Foundations of the Social Sciences: Analyzing Controversies in Social Research*. Cambridge: Cambridge University Press.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Laudan, Larry. 1996. *Beyond Positivism and Relativism: Theory, Method, and Evidence*. Boulder: Westview Press.
- Mahoney, James. 2008. “Toward a Unified Theory of Causality.” *Comparative Political Studies* 41:4–5 (April), 412–436.
- Mayo, Deborah G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.
- Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
- Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge: Cambridge University Press.
- Reichenbach, Hans. 1938. *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University of Chicago Press.
- Rubin, Donald B. 2008. “For Objective Causal Inference, Design Trumps Analysis.” *Annals of Applied Statistics* 2:3, 808–840.
- Seawright, Jason. 2010. “Regression-Based Inference: A Case Study in Failed Causal Assessment.” In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd ed. Henry E. Brady and David Collier, eds. (Lanham, MD: Rowman & Littlefield), 247–271.

Symposium: Research Openness in the Digital Age

Qualitative Data Access and Research Transparency

Colin Elman

Syracuse University
celman@maxwell.syr.edu

“...if people don’t know what you’re doing, they don’t know what you’re doing wrong.”¹

This brief essay draws on recent conversations about data access and research transparency. It discusses some of the issues involved, and describes a vocabulary to handle them. Finally, it explores some of the challenges of increasing openness in the context of qualitative research.

Data Dialogues

Issues of data and research transparency have long been considered important, particularly among political scientists who use quantitative methods. Recently, scholars have revisited a series of linked transparency issues—and expanded the conversations to include scholars who use qualitative techniques.² What responsibilities do scholars have to describe how they collected data and how they analyzed evidence to arrive at their conclusions? How (if at all) does the availability to subsequent scholars of the data and techniques scholars use impact the persuasiveness of the original scholars’ arguments? What is the value of allowing scholars other than those who collected the data to use them for analytic tasks beyond those of the original study?

Members of different research communities in the discipline arrived at these recent conversations with different perspectives. Quantitatively oriented scholars, who have been considering these issues for a longer period, commonly hold a commitment to replication and use data that are relatively less

problematic to share. Correspondingly, their concerns were to clarify these commitments, to consider improvements in infrastructure to facilitate data sharing, and to develop stricter standards, and in the meantime, ensure higher rates of compliance with already widely shared norms.

Qualitative scholars are for the most part both less well-equipped and less willing to join this dialogue. Yet the pressure for qualitative scholars to engage questions about research openness is increasing. Cognate disciplines in the United States like anthropology have made important progress on these issues, and European (and especially British) research communities have pushed further still. In addition, more and more data are being “born digital,” and younger generations of scholars arrive in the academy with expectations that information should be immediately accessible. Funding agencies and journals are also asking qualitative scholars to take a more considered approach to when qualitative data should be made accessible.

Political science as a discipline is constituted by a diverse set of research communities, with different views of the enterprise in which they are engaged. This diversity could easily derail progress, as each of the stakeholders seeks to frame the dialogue in terms that are most sympathetic to their research communities. A more promising approach might be to move forward with a two-stage approach: First, describe general concerns relevant for all types of social inquiry in the discipline, and second, consider how responses to those concerns might be instantiated in the context of research communities with different epistemological commitments.

Developing a Vocabulary to Discuss Data Access and Research Transparency

One challenge in generating an initial “whole discipline” dialogue is arriving at a vocabulary that is relevant to social inquiry in general *and* can accommodate alternate operationalizations for different research traditions. On the one hand, even

at this general level the central terms should have purchase on increasing data access and research transparency. On the other, the terms need to be sufficiently broad to have relevance to different research communities, and thus have the flexibility to be instantiated in various types of social inquiry.

A helpful point of departure is to acknowledge that empirical social inquiry analyzes information about the world to generate intersubjective knowledge statements. While scholarly communities have different purposes and techniques, in the end all strive for the same overarching goal: to produce and present the most *persuasive* accounts possible, within the terms set by their epistemological commitments. A scholar writes for an audience, most often members of her immediate academic community.

A core claim made by those who favor research openness is that, *ceteris paribus*, *transparent* evidence-based research is more persuasive. Scholars try to convince others that their research advances a valid argument. The more information scholars provide about the evidence on which they base their claims and the analytical machinery they employed to arrive at their conclusions, the more swayed we should be. We might consider three dimensions along which the openness of empirical research can be evaluated.³

Are evidence-based knowledge claims backed up with citations to the data they rely upon, including information about where those data were found? If scholars collected their own data (whether via interviews, archival research, or any other data collection technique) do they provide *data access*? Do they let others see their data (be they interview transcripts, primary documents, etc.) or explain why they cannot?

If scholars collected the data on which they rely, do they engage in *production transparency*—do they offer an account of the procedures used to collect or generate the data?

Are evidence-based knowledge claims marked by *analytic transparency*, that is, do they strive to describe the process through which scholars drew their conclusions based on their data?

These questions can apply, *mutatis mutandis*, to all empirical social inquiry. They can be mapped on to different research traditions without forcing a false equivalence among them. In short, we can agree on the general principles underlying these dimensions while acknowledging that they will be instantiated differently depending on the style of social inquiry being conducted.

The Challenges to Openness in Qualitative Research

Qualitative scholars might have several reasonable concerns about research openness. An initial set is listed below, together with some preliminary thoughts about their probable impact.

Epistemologically, many qualitative scholars identify transparency with replication, and believe that repetition by a subsequent scholar for the purposes of assessing validity is

neither possible nor desirable.

An inclusive dialogue about openness should be sensitive to these types of epistemological concerns, and explicitly eschew imposing inappropriate norms on types of social inquiry where they do not belong. While transparency is a necessary condition for replication, the opposite is emphatically not so. To say that providing more information about a knowledge statement strengthens one's argument does *not* imply that another scholar armed with the same information must be able to arrive at the same conclusion in order for that argument to be considered valid. In this context, the case for transparency is that the more material scholars include, the more sense readers can make of their arguments and the more convinced they may be by them. As one scholar observed, a review essay on a Shakespeare sonnet is likely to be better understood and more persuasive if readers also get to read the sonnet.

Of course, some scholars who use interpretive methods are comfortable with at least some types of replication (see, for example, Wedeen 2010: 265). Moreover, many members of the qualitative and multi-method community who aim to produce objective social science are at least notionally committed to the idea that scholars using the same research tools in the same sites will collect similar data and arrive at similar conclusions and arguments. Consider, for example, the "positivist" qualitative dissertation which goes on to be published as a book by an academic press. Archetypically, the volume comprises six chapters: an introduction, a discussion of relevant theory and the study's research design, three case study chapters based on data gathered through interactive or archival techniques, and a conclusion. Although rarely put to a direct replication test, work of this type does implicitly claim to convey objective information and draw conclusions that other scholars employing the same research techniques would convey and draw. Yet while there is an abstract assurance of openness in this type of qualitative research, for the most part research transparency amounts to a few paragraphs in the text discussing where the author found her data and how she collected them. If subsequent scholars want to actively evaluate her claims, they have to pay the transaction costs of visiting the data sites themselves.

The point here is not to re-litigate the arguments for and against replication. It is to argue that research openness is a broader ideal, and one from which scholars can benefit regardless of which viewpoint they take on replication. For qualitative scholars who abjure replication, transparency strengthens the persuasiveness of their inquiries and their arguments. For researchers who believe that a necessary element of a truth claim is that other scholars should be able to duplicate conclusions, transparency is a prerequisite.

Pragmatically, researchers worry that committing to transparency will make data collection (be it through interaction or archival research) more difficult, with limited pay-offs for the additional work.

Valuing openness does not require blindly following the principle to the detriment of research. Transparency is one among several competing objectives researchers might pursue, and it will not always take priority. For example, field re-

searchers do not expect to be able to videotape or record every one (and sometimes any) of their interviews. In some instances contemporaneous notes (or simply listening!) will be the best they can do. Researchers' main responsibility is to get the best data they can, using the approach most appropriate to the circumstances. Where factors prohibit researchers from sharing their data, or require them to present it in a particular form, they simply need to offer a clear account of what those obstacles were. With respect to payoffs, any reasonable conversation about research transparency and data access has to include the establishment of guarantees of first-use of data by their collectors/generators, but also the introduction of disciplinary reward structures that incentivize scholars to take on the burdens imposed by transparency.

To be sure, these types of pragmatic concerns are real and deserving of serious attention. But the absence of an American tradition of openness in qualitative social inquiry may also have generated a tendency to dwell on difficulties. Some of these concerns are likely to seem less worrisome once more public examples of transparent qualitative research are presented. In addition, because openness is achievable by a range of possible alternatives, once some of these pathways to transparency are described and used, more are likely to be developed. With a wider range of different approaches, it is likely that some may be better suited to address particular pragmatic concerns. This is another reason why it is better to think of the endpoint of this conversation as a menu, not a mandate.

Of course research openness will likely involve its own pitfalls and unintended consequences. For example, at the moment, scholars whose main data sources are primary documents cite, but do not provide links or access to, the primary sources underlying their claims. By accessing an author's primary sources by visiting an archive, a second researcher can both investigate whether those sources really do support the argument, *and* check for other sources that undermine it. Once authors begin to provide direct access to the primary sources they cite (or at the very least, substantial redactions), checking referenced materials will become much easier than exploring what materials might have been ignored. While this is not a good argument against openness, it is a strong reminder that transparency is a matter of degree and one step in a broader evaluative process. We should not confuse a scholar's provision of materials that underlie her claims with undeniable proof that those claims are valid, nor make the assumption that she has provided every piece of evidence that might bear on the argument.

Ethically, qualitative scholars quite properly worry that providing others access to their data may involve human subjects or legal concerns.

No conversation about openness in qualitative research can proceed without carefully considering ethical and legal obstacles to transparency, and including those concerns as valid grounds for withholding information. With regard to finding areas on which scholars can agree, this is the most straightforward of all the obstacles. Protection of human subjects, for example, is a concern which unites all researchers. The rule should be correspondingly straightforward: if the protection

of human subjects constrains what scholars can safely say, they should work within those limits. But they should also explain to readers why the data are unavailable, and (depending on the epistemic community in which they are operating) perhaps consider making the data available in an anonymized form.

Conclusion

Increasing openness is likely to have positive effects on qualitative research in the long run. As data become more immediately visible, and as the bar is raised for requiring explicit connections between authors' data, analysis, and conclusions (i.e., as they are encouraged to demonstrate how particular techniques were used to make inferences from particular data), we anticipate research designs and execution will become more rigorous.

In addition, there will be substantial pedagogical opportunities from being able to demonstrate the clear application of analytical techniques. To be clear, this does *not* require or even encourage convergence on a single method of data collection or data analysis. But it will need scholars to be much clearer about the choices they make among the alternatives.

We understand that there is some reticence about openness among different groups of qualitative researchers, but we nevertheless believe that this is a conversation worth having. Otherwise the discipline will push ahead without the active participation of qualitative researchers. We will either be left behind entirely or, worse, be squeezed into an unhelpful one-size-fits-all template. In addition, the wider discipline's view of qualitative research is likely to be heavily shaped by technological possibilities rather than epistemological imperatives. The worst case scenario is that qualitative research will be the passive ingredients to big data's synthesis of combine harvester and food blender, indiscriminately harvested and homogenized. A better solution is to instead begin with the research traditions, and develop transparency tools that fit.

Broad participation in shaping the dialogue about openness will maximize its benefit across the discipline, and allow us to shape application of transparency norms to fit our research styles. Accordingly, this short memorandum argues that qualitative researchers will be better off if they have a positive agenda that is the product of a conscious conversation, grounded in our epistemological, pragmatic, and ethical concerns.

Notes

¹ Lynn and Jay (1984: 21). Jonathan Lynn and Anthony Jay authored the BBC series *Yes Minister*, and wrote a book drawn from the series in the form of Jim Hacker's fictional diaries (but actually comprising a combination of those diaries with fictional memoranda and notes by other participants). The quotation in the epigram is from a note by Humphrey Appleby, and attributes the observation to Sir Arnold Robinson, Secretary to the Cabinet.

² In August 2010 Arthur Lupia of the University of Michigan and I co-authored a memorandum to the APSA Council, which resulted in the formation of an APSA Working Group on the topic of research transparency. Although the current essay draws on and reflects several themes of that dialogue, it does not constitute a report from the

Working Group or purport to represent the views of any of its other members. The DA-RT Working Group's recommendations were reviewed by the APSA Governing Council in September 2011, and passed to the Ethics, Rights, and Freedom Committee. The text was returned to Council in April 2012, and the language adopted as APSA policy.

³ These terms were originally introduced into the current discussions in the Lupia and Elman memorandum referenced in note 2 above.

References

- Lynn, Jonathan and Antony Jay. 1984. *The Complete Yes Minister: The Diaries of a Cabinet Minister, by the Right Hon. James Hacker MP*. Revised omnibus edition, London: British Broadcasting Corporation.
- Wedeen, Lisa. 2010. "Reflections on Ethnographic Work in Political Science." *Annual Review of Political Science* 13, 255–272.

Closing the Infrastructure Gap: Qualitative Data Archiving

Diana Kapiszewski

University of California, Irvine
dianakap@uci.edu

Over the last 15 years, political science has witnessed a renaissance in qualitative research methods (see, e.g., Brady and Collier, eds., 2010).¹ The canon has been reworked, new areas of scholarship have appeared, and a rapidly expanding body of political science research now employs qualitative and multimethod analysis. Correspondingly, as noted in this symposium's introductory essay on the openness dialogue, although qualitative researchers have begun to explore ways to share their data and access those of other scholars, the lack of a dedicated venue or consensual set of practices for storing, sharing, and reusing qualitative social science data in the United States (Heaton 2004: 6) presents a significant obstacle. This infrastructure gap—which transcends scholarly differences over the contributions of qualitative research—contrasts sharply with well-established norms in quantitative research, and with the practices of qualitative social scientists in other countries.² As a result of the lack of an appropriate data-sharing venue, the few American social scientists who *do* share their qualitative data generally do so via inefficient ad hoc arrangements.³

This brief essay argues for the development of generalized norms and specific practices for archiving and sharing qualitative data, and discusses an ongoing initiative to create a dedicated qualitative data repository. Archiving and sharing qualitative data will ease evaluation and replication of research, render research processes more transparent, and encourage secondary data analysis. Doing so will also provide valuable pedagogical tools and, by increasing researcher visibility, promote the formation of epistemic communities and research partnerships. Of course, not all qualitative data are shareable, and the establishment of norms of sharing could have unintended

and sometimes negative consequences. Nevertheless, the potential rewards of qualitative data archiving arguably compensate for the efforts required to address its difficulties.

The Promise of Qualitative Data Archiving

Qualitative data archiving enables scholars to store, search, access, and download electronic qualitative data of all types, from official documents, to interview transcripts, to photographic, audio, and video materials. Archiving qualitative data can produce several important benefits.

First, qualitative data archiving will allow for the vertical integration of primary data, secondary analysis, and scholarly output, allowing scholars to provide access to the data they used to arrive at their inferences and interpretations, and thus better demonstrate how they developed them. This transparency will encourage researchers to carry out data collection and analysis in a systematic, replicable way. It will also allow scholars to learn from others' experiences, help them to avoid reproducing mistakes, and facilitate discussion and critique of qualitative methods. Second and relatedly, by making data available and increasing the transparency and visibility of research processes, qualitative data archiving can dramatically reduce the costs of assessing and replicating empirically based qualitative analysis (Swan and Brown 2008: 7).

For instance, qualitative data archiving will facilitate instantiation of the "active citation" standard advocated by Andrew Moravcsik in his contribution to this symposium and elsewhere (e.g., 2010) by mediating between scholarly references and hyperlinked sources. This specific type of data archiving will allow scholars to make timely comments on and corrections to other scholars' use of primary sources. Consider, for example, the erroneous citation of a document as diagnostic evidence in the context of a process tracing narrative. Under the present state of affairs, the mistake would likely go unnoticed absent a subsequent publication on a closely related topic. The primary document's posting to a qualitative data archive would permit more immediate feedback at much lower transaction costs.

Third, archiving qualitative data will provide valuable pedagogical and coordination tools. Students taking qualitative methods courses will be able to learn from and critique the data-collection techniques used by scholars who archived their data, better understand the analytic strategies such scholars used in their published work, and practice the analytic techniques they are learning on real empirical data. Also, the publication of data will vastly increase the visibility of scholars working on particular topics, facilitating team research and the formation of epistemic communities around research areas and questions (Swan and Brown 2008: 26). Finally, archiving qualitative data will facilitate data accumulation, allowing scholars to undertake research in the context of a much larger universe of available data, and to make comparisons across space, time, policy areas, groups, and so on that could otherwise require additional research resources or assembling a research team (Corti 2000: 6.2).

One concern might be that expanding the practice of qualitative data archiving could weaken the current norm (or even

obligation) among qualitative researchers of close engagement with cases. At the limit, one could imagine qualitative research being undertaken without scholars directly engaging with *any* of the cases from which the data under analysis were drawn. Yet it seems very unlikely that qualitative researchers would become so disconnected from their cases either in the short- or long-term, not least because doing so would undermine one of the primary comparative advantages of qualitative research.

Given the potential benefits of archiving qualitative data, several scholars have initiated a project to develop a dedicated digital qualitative data repository.⁴ The repository will be accessible via a website portal, with user-access controls to regulate how scholars search and retrieve archived data. This archive will of course be just one place qualitative researchers might store their digital data. They can also store them in general archives like Dataverse and ICPSR; in university repositories such as those offered by Cornell, Duke, the University of California, and many others; in archives focused on particular issue areas; and on scholars' own personal websites. The main goals of the dedicated qualitative data repository currently under construction—which will aim to link with these other venues—are to demonstrate the practical possibility and intellectual promise of sharing qualitative data broadly, encourage their sharing, and serve as a site around which best practices can begin to be developed.

Strategies for Addressing the Challenges of Qualitative Data Archiving

Qualitative data sharing presents a range of challenges—and requires a set of solutions—that differ in some ways from those associated with quantitative data archiving. A first set of issues concerns data-collection practices. Making research procedures more transparent may have the unintended consequence of encouraging researchers to engage in self-censorship, for instance, omitting from their analyses data collected using a technique they fear will not be considered rigorous. A related problem is that if the “shareability” of the underlying data becomes an important criterion for judging empirical qualitative research, scholars may focus solely on contexts where they can collect data that can be shared easily. Important information that could have been collected and used with discretion will go unsolicited, and important topics will not be researched.

A connected series of concerns regards the particular dynamics that might limit the sharing of qualitative data. Institutional review boards (IRBs) are likely to require that researchers who wish to store and share data collected using interactive techniques such as interviews and focus groups solicit subjects' permission (Mauthner, Parry, and Backett-Milburn 1998: 743; Heaton 2004: 79). Inevitably, some subjects will be unable or unwilling to have their identity revealed and/or information they provide made available to other scholars (Corti, Foster, and Thompson 1995: 3). Further, interviews are sometimes given off-the-record and for background purposes only, and subjects are sometimes promised anonymity (Heaton 2004: 81; Parry and Mauthner 2004: 146). On the one hand, the conditions under which subjects offer information can affect the

type and form of data scholars collect. On the other, some data may require a resource-intensive process of contextualization in order to be shared (or may need to be stored in partial form). All of these eventualities can have significant consequences for analysis, interpretation, and inference.

Protecting human subjects and conforming fully to standards set by IRBs are crucial imperatives. Achieving those goals may prevent some data from being shared. Yet a range of measures can be taken to protect human subjects while sharing data collected from and about them. For example, washing qualitative data can help preserve anonymity; scholars who need to edit their data in this way can explicitly note what sort of information was removed and assess the impact of its deletion on the remaining data. Differential user access to digital archives can also help to address human-subjects concerns.⁵ In sum, these important concerns need not stand in the way of making a great deal of material available to a broad range of scholars.

Non-interactive forms of data collection, such as archival research, may also produce data that cannot be electronically archived, or that can only be posted with some time lag or in attenuated form. For example, documents accessed in archives may have copyright restrictions that prevent them from being reproduced or made publicly available for secondary reproduction. Like those related to human subjects, these challenges require specialized solutions that it is very likely social scientists can develop through sustained, thoughtful debate and collaboration.

In sum, while not all data social science researchers collect and produce as part of qualitative and multi-method research can be archived and shared, a considerable amount likely can. Sharing a significant subset of the qualitative data with which social scientists work is vastly superior to the status quo, in which practically no qualitative data are publicly available in the United States.

The Road Ahead

Disciplinary norms will need to change if sharing and re-using qualitative data—and producing scholarship relying on secondary data analysis—are to become accepted practices. As noted in this symposium's introductory essay on the openness dialogue, the discipline could benefit from a sustained conversation about the merits and limitations of data sharing, and from clarifying guidance on data release and sharing in qualitative research.⁶ Likewise, funding agencies could advise grantees to make qualitative data available, and more journals could require authors to do so.⁷ Training opportunities, outreach, and guidance on qualitative data sharing (see Corti 2000, 6.3) could also be offered through the Institute for Qualitative and Multi-Method Research or the short courses taught at the annual APSA meetings.

Establishing a dedicated qualitative data repository like the one currently being developed, rather than building on a pre-existing quantitative archive, may be the most effective way to encourage qualitative researchers to deposit and use shared data. Such a specialized repository will be more attuned to the particular challenges posed by archiving qualitative data

References

and will have a larger demonstration effect. The repository will also be epistemically neutral—viewed broadly as a means of increasing the transparency of the evidentiary basis for interpretive, descriptive, or explanatory work based on qualitative data—and designed to be visible to, and open to communication and interaction with, a wide audience. And of course, as an electronic resource, the repository will be linked to the broad range of existing institution-specific and specialized archives that already exist.

Qualitative research makes vital contributions to political science, and qualitative data archiving holds the key to making qualitative and multi-method research more transparent and more replicable. Moreover, sharing allows data to be used as a basis for further research, and encourages scholars to engage in secondary data analysis, opening up a range of new research possibilities, including cross-temporal and cross-context comparison. Of course, as occurs whenever new practices may be adopted, the challenges and risks of sharing and reusing qualitative data must be carefully considered and addressed. Nonetheless, those challenges may prove to be relatively minor in comparison with the tremendous utility that sharing and reusing qualitative data can provide.

Notes

¹ This piece draws extensively on an article that appeared in the January 2010 issue of *PS*, co-authored with Colin Elman and Lorena Vinuela. I would also like to thank the broader set of scholars who participated in a workshop convened to explore the idea of building a qualitative data repository held March 28–29, 2009, at Syracuse University (funded by NSF Grant SES 0838716).

² For instance, funding agencies in several OECD countries adopted a mandatory sharing policy for grant holders in the 1990s, and the repositories constructed as a result receive regular deposits on a national scale and hold a wide range of qualitative materials. Some examples include QUALIDATA in the UK, WISDOM in Austria, SDA of the Czech Republic, DDA of Denmark, FSD in Finland, Réseau Quetelet in France, GSDB-EKKE in Greece, GESIS in Germany, ADPSS Sociodata in Italy, CEPS in Luxembourg, DANS in the Netherlands, NSD in Norway, ARCIS/CIS in Spain, and SND in Sweden.

³ To be sure, several university libraries and research institutions have archives for data collected by their affiliated researchers and the facilities to archive digitalized text and audio material. Nevertheless, in the American academy the overwhelming focus is on archiving quantitative data, or on quantitative redactions of qualitative data.

⁴ The project has been funded by NSF Grant SES 1061292.

⁵ For example, some data may be made available for online use by any registered scholar while other data may be kept in non-networked storage to be accessed only in person at the repository with the depositor's (and if need be, the original source's) permission and in accordance with explicit data-sharing agreements.

⁶ The American Political Science Association's *A Guide to Professional Ethics in Political Science* (2008) envisions non-release as the default—except when funder mandates or challenges to findings trigger release. Although the guide establishes a general heuristic requirement to disclose non-confidential sources for replication and testing, it does not specify whether “sources” refers to the identity of interviewees or to data.

⁷ Most political science journals that have data-release policies either explicitly or implicitly limit those mandates to statistical data.

- American Political Science Association. 2008. *A Guide to Professional Ethics in Political Science*, 2nd ed. Washington: American Political Science Association.
- Brady, Henry and David Collier, eds. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd ed. Lanham, MD: Rowman & Littlefield.
- Corti, Louise. 2000. “Progress and Problems of Preserving and Providing Access to Qualitative Data for Social Research—The International Picture of an Emerging Culture.” *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 1 (December), <http://www.qualitative-research.net/index.php/fqs/article/view/1019>.
- Corti, Louise, Janet Foster, and Paul Thompson. 1995. “Archiving Qualitative Research Data.” *Social Research Update* 10 (Autumn), <http://sru.soc.surrey.ac.uk/SRU10.html>.
- Elman, Colin, Diana Kapiszewski, and Lorena Vinuela. 2010. “Qualitative Data Archiving: Rewards and Challenges.” *PS: Political Science & Politics* 43:1 (January), 23–27.
- Heaton, Janet. 2004. *Reworking Qualitative Data*. London: Sage Publications.
- Mauthner, Natasha S., Odette Parry, and Kathryn Backett-Milburn. 1998. “The Data Are Out There, or Are They? Implications for Archiving and Revisiting Qualitative Data.” *Sociology* 32 (November), 733–745.
- Moravcsik, Andrew. 2010. “Active Citation: A Precondition for Replicable Qualitative Research.” *PS: Political Science & Politics* 43:1 (January), 29–35.
- Parry, Odette and Natasha S. Mauthner. 2004. “Whose Data Are They Anyway? Practical, Legal and Ethical Issues in Archiving Qualitative Research Data.” *Sociology* 38 (February): 139–152.
- Swan, Alma and Sheridan Brown. 2008. “To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs.” A report commissioned by the Research Information Network. School of Electronics & Computer Science, University of Southampton. <http://www.rin.ac.uk/files/Data%20publication%20report.%20main%20-%20final.pdf>.

Active Citation and Qualitative Political Science

Andrew Moravcsik
Princeton University
amoravcs@princeton.edu

This article presents a proposal for the adoption of “active citation,” together with a discussion of why it is necessary, its possible advantages, and some potential concerns.¹ Active citation envisages the use of rigorous, annotated citations hyperlinked to the sources themselves. The goal is to provide opportunities for scholars to be rewarded not just for more rigorous but also for richer and more diverse qualitative scholarship.

The Problem: The Evidence in Qualitative Research Remains Invisible

Qualitative research dominates political science. While the use of statistical and formal methods is spreading, historical, qualitative, or textual research remains strong. In the field

of international relations, for example, roughly 70% of scholars still primarily conduct qualitative research, compared with 21% chiefly favoring formal or statistical analysis.² Hardly any major political science debate remains untouched by important qualitative contributions. Yet this underestimates the impact. Since nearly all quantitative scholars make secondary use of textual methods, overall *over 90% of scholars* employ qualitative analysis, whereas only 48% use any statistical and only 12% any formal methods. And still this understates the impact, because many statistical data sets rest ultimately on historical and textual analysis.³ Were that not enough, when we look to academia's impact on the world, qualitative case studies are reported to be more relevant for policy than quantitative or formal work.⁴

Despite the importance of qualitative case studies to political science, the textual evidence on which they rest remains largely invisible.⁵ To be sure, footnoting practices formally require that authors specify where they found documentary support for controversial empirical claims. Yet the resulting citations would only frustrate most readers who seek to track the evidence. Some scholars may simply seek to trace the causal process in order to better understand it. Others may want to criticize or challenge the argument. Still others may wish to know whether well-known standards of unbiased qualitative causal inference (e.g., process-tracing, case selection, primary source selection, etc.) were adhered to. Some may seek to improve and supersede the empirical findings. Others may seek to do secondary analysis, pooling the data into larger sets. Access to the textual evidence may serve unrelated scholarly purposes. More interpretive scholars less concerned with causality may seek more direct and unmediated appreciation for the subjective experiences and the plurality of voices in the past and present.

Nearly all scholars, no matter what their preferred method or epistemology, believe that the ability of scholars to engage in the activities above is essential to healthy scholarship. Yet none is likely to be possible with contemporary qualitative political science, because of the way it presents and manipulates textual evidence. This is so for three basic reasons.

First, citations in political science are sometimes frustratingly imprecise. It is not uncommon for scholars in political science to back a claim with a simple citation to an article, chapter, or book, without page numbers or other specific reference. This would not be permitted in other, more textually self-conscious disciplines, since it effectively precludes the reader from linking concrete evidence to a general claim.

Second, citations often lack a specific quotation or annotation illustrating exactly how and why the citation supports the textual point. This is in part due to the fact that political science has never developed such practices, which are common in academic history and mandatory in legal academia. This tendency toward vagueness has been exacerbated recently in articles and books by ever tighter word limits and the spread of "scientific" citation forms designed for fields in which every citation is to another secondary social scientific article. These are manifestly inappropriate for qualitative research, because they preclude minimally rigorous explanation of pri-

mary textual evidence and its relationship to interpretations of events.

Third and even more important, even when political science citations are precise and elaborated, readers often find it prohibitively expensive (in time and money) to view the evidence underlying interpretations and empirical claims. In theory, of course, the reader can simply "check the source." Yet, in practice, this is possible only a small percentage of the time.⁶ If the reader is lucky, the claim can be checked using an online source, such as an electronic newspaper and journal collection, government document archive, or a secondary source reproduced in Google Scholar. Yet original points in almost all major works of empirical consequence rest primarily on other sources. At best, such sources might be published secondary books. Locating and reading these is likely to require that the reader access numerous libraries. For international work, this could well be in several jurisdictions, various countries or more than one language. More likely, serious qualitative research in political science rests, to a substantial extent, on collections of informal publications, archival material, interviews, participant-observation or ethnographic observation notes, that are available only locally. These may be subject to human subject and proprietary concerns, may also be in a foreign language, or make take the form of notes, tapes, scans, or photocopies.

Any scholar who seeks to understand, replicate, criticize, or build on an article or book by following its evidence is, more often than not, effectively precluded from doing so. Simply reassembling the sources cited in an article about politics could well require a research commitment that resembles in scale that of the original author. Alternatively, one could ask the author for the evidence—but it is unclear why or how most authors could comply with such requests. The result: It is extremely rare for the quality or veracity of textual evidence, or its connection to argument, to be challenged, let alone for political scientists to replicate, improve, extend, or reuse qualitative evidence.

This stifles opportunities for debate, diversity, and progress in qualitative political science. To see why, compare this to the situation prevailing in quantitative (statistical) political science. Much work rests on publicly available datasets, and a norm exists whereby new datasets are made public. To be sure, many scholars believe quantitative work might also benefit by improving transparency. Yet it has advanced far enough that findings can and are often replicated and extended, and data is used for secondary purposes. Indeed, a common first-semester exercise in political science statistics courses is to replicate a major article and then extend it by adding data from another source. Such an exercise is unheard of in historical and qualitative areas of political science. Indeed, because of its lack of transparency, unclear standards, and resulting insulation from challenge, qualitative political science (despite its lower technical and training demands) has gained an aura of aloofness and elitism among many graduate students.

Good reasons exist to believe that the quality, rigor, richness and future attractiveness of non-quantitative political science within political science rests in large part on improving the management of textual evidence in such a way as to make

it more easily and broadly available to the research community of qualitative scholars.

The Proposal: Active Citation

One way to overcome the prohibitive cost of accessing qualitative information is to establish a universal standard that assures transparency and replicability in selection, presentation, and preservation of textual evidence. The standard proposed here is *active citation*: the use of *rigorous, annotated citations hyperlinked to the sources themselves*. This proposal seeks to exploit new technologies to generalize to political science the best practices in history, law, sociology, and the natural sciences with regard to the presentation of evidence. The proposal rests on three general principles.

1. Precision: Any critical and contested substantive empirical claim in a qualitative case study must be backed by a citation to one or more concrete sources. The citation should be precise enough to unambiguously identify the page(s) and passage(s) of the source that backs the claim. Such sources may be unpublished primary sources, published primary sources, primary sources cited in secondary sources, secondary sources, research materials, or other evidence. While normally textual sources are envisaged, there is no reason—modern technology being as it is—why visual, graphical, photographic, audio and other materials could not be inserted.

2. Annotation: The citation should be annotated to explain precisely how the source supports the textual claim, and informing the reader of any contextual information essential to an interpretation of the source text.

3. Transparency: Each citation to a controversial empirical claim would contain a hypertext link (within the document) to a reproduction or transcript of the source material in context. This will appear (in order of citation) in an Appendix, which would appear in electronic versions of journal articles, in unpublished papers, and in parallel electronic versions of the notes of scholarly books that only appear in hard copy. Normally this contextual source material will comprise a presumptive minimum one to two paragraphs, or around 100 words, but this total may vary by circumstance.

How would active citation appear in practice? The footnote, endnote, or in-text note would contain a precise citation. It would appear much as citations do currently, though there would be a firm expectation that it be complete. On the hard-copy version of most political science articles, where that is all one sees, the article would thus appear unchanged from what appears today. On an electronic version, however, that citation would be hyperlinked to an excerpt from the corresponding source, which would be found in an appendix to the document, listed with other sources in order of citation. The third element, the annotation, could appear in the appendix with (immediately preceding) the corresponding source material, or in the footnote, if that is compatible with the format of the journal or book. The latter is preferable, common in historical journals, and almost universal in legal journals—but journal and book editors would be free to retain current practices in this regard.⁷

Whatever the precise formatting choices, the common result would be that, when reading the electronic version, a reader could:

(a) access and read the source in real time with one click, and return to the text with one click;

(b) procure a complete list of all the cited sources (on controversial empirical points), in order of citation, by simply copying the appendix.

This format can be approximated by manipulating existing commands in commonly-employed word processing programs (e.g., Word, Latex), but within a relatively short period a team of us working on the issue expect to develop specific software that will streamline the process.

By adopting active citation, political science would approximate the standards normally expected in academic disciplines where textual analysis and interpretation, and research transparency, are more refined and rigorous than in political science. Legal scholars who publish in law reviews and journals are accustomed to reading articles with precise and complete footnotes, with annotation and quotations to specify why they support the claim, and hyperlinks to any available sources.⁸ One difference here—largely for intellectual property and human subject protection reasons, as well as logistical ones—is that the presumptive standard will be that linked sources are excerpted and included in an appendix to the article, rather than cited extensively, included in long footnotes, or left as hyperlinks. In history, similarly, footnotes are precise and often more extensively annotated, including quotations. By actively employing the appendix to reveal data and analysis, the criteria proposed here also resemble standard practice in natural science journals today, where relatively short articles are followed by often extensive “Supplementary Materials” sections containing data, experimental results, charts, further analysis, background, video, spreadsheets, etc.⁹

The Advantages: Promoting and Rewarding Qualitative Research

Active citation promises to improve qualitative political science in four main ways. Most obviously, it would encourage and reward higher-quality scholarship. Researchers who face an immediate requirement of precise, annotated transparency will be motivated to carry out data collection and analysis in a more careful, systematic, and replicable way, and to report the precise empirical basis of empirical claims they advance. Greater transparency and replicability would help unleash the full potential of analytic narratives, fine-grained process-tracing methods, and strategic case selection. Scholars will face greater incentives to improve their qualitative methodological skills. When proper adherence to such methodological standards becomes a transparent act that others can observe and evaluate, expert use of the method and superior qualitative data collection can be properly recognized and rewarded within the profession—which does not occur often today. In addition, virtues such the ability to read texts carefully and creatively, to place them in historical and cultural

context, to speak and read foreign languages, and to appreciate multiple perspectives, may well increase in importance.

While active citation encourages more careful research, it will also empower critics. By revealing evidence at a single click, active citation would democratize the field, letting new and critical voices be heard. Any potential critic could make an immediate assessment of the evidence for empirical claims and its relationship to the research design, theory, and method. A graduate student anywhere in the world would require but an afternoon to decide if a published qualitative argument is *prima facie* plausible. Flaws like selective citation, poor use of sources, or contextually inappropriate interpretation would become far easier to document. Livelier and more engaged scholarly debate—in the form of criticism, replication (“research auditing”), and review essays—would be encouraged.

Active citation would likely also encourage more “secondary” analysis of qualitative evidence, that is, the use of textual evidence for alternative purposes. Today most qualitative political scientists start essentially from scratch.¹⁰ This stands in striking contrast to quantitative studies, where each scholar can build on previous data-collection efforts, and the pool of data expands over time. Active citation would erase this imbalance. Much of the evidence of existing scholars, in electronic form, would already be available to fresh students and scholars entering a field, who would have a greater incentive to collect new evidence—since secondary analysis would only need to provide a *marginal* increment of new evidence to make an original contribution. The mark of a healthy scientific research tradition is precisely that over time debate encourages *ever increasing amounts of data* to be revealed in this way.¹¹ The expanding network of available data could also facilitate meta-analysis, in which comparative analysis of similar situations in various settings (e.g., countries, issues, time periods) could be conducted using different evidence/data, perhaps with addition of new sources provided by the investigator. Declining start-up costs for each new scholar working on a topic would encourage new scholars to join the “club” by contributing new data, just as combining pre-existing statistical data sets with new data reduces the costs of doing quantitative work on existing topics.

Finally, active citation may well encourage more intensive interdisciplinary interaction. Placing qualitative political science on a more transparent foundation would open debates to a wider range of voices, interpretations, and perspectives, including an expansion of opportunities for interdisciplinary interaction. Certain branches of law, history, and sociology, we have seen, employ higher qualitative research standards with regard to citing, documenting, and presenting evidence than those that currently prevail in political science. Creating incentives for political scientists to engage legal and historical scholars may encourage the formation of an interdisciplinary critical mass employing similar methods, standards, and evidence.

The Concerns: A Word to Potential Critics

Some may worry that active citation places an excessive logistical burden on scholars, journal editors, or publishers; that it is inconsistent with appropriate intellectual property

rights for scholars to exploit evidence they have uncovered; that it would lead to free riding by those who do not actually collect evidence. I have answered these criticisms elsewhere.¹² For the most part, I find them without much substance—in part because other disciplines have adopted similar practices without ill consequences. Where there are legitimate costs and concerns involved in adopting this proposal, they are almost certainly outweighed by the individual and collective advantages for qualitative scholars of all types, and the profession as a whole.

In an earlier proposal, I took special care to engage colleagues concerned that active citation may, by its very nature, encourage research that does violence to the specific historical, ideological, cultural, personal, and gendered context in which researchers and research subjects interact. Some among historians, ethnographers, and non-positivist political scientists view the interpretation of sources as a fundamentally reflexive or hermeneutical process, with much “local knowledge,” contextual understanding, deep expertise, creativity, and hermeneutical interaction required to grasp meaning. They may believe that the proliferation of one to two paragraph snippets, even with annotation, would encourage a superficial, even positivistic, understanding of textual evidence. For this reason, one can imagine some preferring that scholarly debates remain restricted to a small number of insiders, as they are today, even with the resulting costs in scholarly quality and anti-democratic hierarchy. Others may believe that the notion of transparency, replication, and rigor that underlie some of the advantages of active citation are too positivistic in spirit for their taste. These are important considerations for many, and they should be taken seriously.¹³

Still, on balance, I believe active citation is likely to constitute an improvement over current practices, not just for those who view qualitative methods as a way of evaluating causal claims, but for those who view such methods as a way to describe, interpret, critique, or engage the past—whether in a traditional historical, area studies, interpretivist, constructivist or post-modern vein. Indeed, such scholars may well benefit the most from a broadening of professional norms foreseen by active citation. This is true in part simply because active citation aims to strengthen qualitative social science across the board, which is the method most such scholars employ. Moreover, active citation promises to validate scholarly virtues of hermeneutical subtlety, contextual understanding, and (inter-)cultural literacy in which such scholars excel—and ultimately expand the community of those who practice them.

In these ways, active citation does not promise simply to improve the rigor of research, but also its richness. It can help expand the range of evidence being considered; increase the number of plural, multiple, and conflicting voices qualitative social science can capture; extend the depth and intensify the immediacy of engagement with new cultural and textual materials; enlarge the contextual variety of descriptions and interpretations scholars can advance; and multiply the voices in the scholarly community empowered to engage in scholarly debates. In all these ways, active citation can be understood as a way of transforming traditional hierarchies of control and

Undergraduate Scope and Methods Courses: Is There Room for Qualitative Methods?

Cameron G. Thies

University of Iowa
cameron-thies@uiowa.edu

publication into an open virtual network, in which new and plural streams of evidence and interpretation, narratives, and debates can emerge—while still imposing discursive rules that require some substantial commitment from serious participants in the scholarly debate, and permit others to voice their objections. Hence all qualitative scholars concerned with texts—whether committed to causality in a conventional sense, traditional historians, and ethnographers, or interpretivist, constructivist, or post-modern in inclination—should welcome such a trend.

Notes

¹ This discussion builds on a previous article. Andrew Moravcsik, “Active Citation: A Precondition for Replicable Qualitative Research,” *PS: Political Science and Politics* 43:1 (January 2010), 29–35.

² Richard Jordan, Daniel Maliniak, Amy Oakes, Susan Peterson, and Michael J. Tierney. 2009. “One Discipline or Many? TRIP Survey of International Relations Faculty in Ten Countries.” Teaching, Research, and International Policy (TRIP) Project, Institute for the Theory and Practice of International Relations, The College of William and Mary, http://irtheoryandpractice.wm.edu/projects/trip/Final_Trip_Report2009.pdf.

³ See, for example, the MIDipedia project led by Michael Tomz and Jessica Weeks, which aims to use historical research to overcome persistent coding problems in the widely-used COW (Correlates of War) dataset.

⁴ Jordan et al. (2009).

⁵ I have written this essay using “textual” evidence as the standard example, which it remains in the field. However, all the arguments apply equally to visual, audio, graphical, and any other multi-media materials that can be stored electronically.

⁶ The exceptions tend to be work within regional studies defined by particular geographical areas. There small communities of scholars exploit local interpretive knowledge, linguistic skills, and a more familiar body of sources, functioning similarly to historians. Another exceptional category contains work in which political scientists recapitulate positions and sources from preexisting historical literature—as in the debate in security studies over the causes of World War I. Keir A. Lieber, “The New History of World War I and What It Means for International Relations Theory.” *International Security* 32:2 (2007), 155–191.

⁷ An example and a protocol can be found on line at www.princeton.edu/~amoravcs.

⁸ See, for example, the *Yale Law Journal* online, accessible through <http://yalelawjournal.org/>.

⁹ See, for example, articles in *Nature* (<http://www.nature.com/>) or *Science* (<http://www.sciencemag.org>).

¹⁰ “Most data generated by American qualitative and multi-method social science are used only once.” Colin Elman, Diana Kapiszewski and Lorena Vinuela, “Qualitative Data Archiving: Rewards and Challenges.” *PS: Political Science and Politics* 43:1 (January 2010), 23.

¹¹ For healthy examples in international relations, against the odds, see, e.g., Jack S. Levy, Thomas J. Christensen, and Marc Trachtenberg, “Correspondence: Mobilization and Inadvertence in the July Crisis.” *International Security* 16:1 (1991), 189–203.

¹² Moravcsik, “Active Citation,” 33–34.

¹³ This is of particular concern to me, since I am trained as a historian and work across several academic disciplines, conduct area studies research requiring local cultural knowledge, have served as a policy-maker in very diverse cultural contexts, and would like to see a broader range of views expressed in academic debates.

The debate over methodology in the discipline has shifted in remarkable ways within the last decade. Qualitative analysis, which used to be perceived as representative of old-fashioned or even empirically sloppy work by its detractors, has now regained a respected role in causal and interpretive political science. Mixed method research that combines formal or statistical models with small-N studies is almost expected of graduate students producing dissertations in many departments. Advances in qualitative methodologies have furthered the goal of identifying causal mechanisms for those pursuing this type of work. The foundations of political inquiry in general have been strengthened by renewed attention to qualitative methods, much of which has been fostered by the Consortium for Qualitative Research Methods (CQRM) and the faculty and graduate students that have come to create such a vibrant community of scholars.

It is very apparent that rigorous qualitative methods courses are increasingly taught in the graduate curriculum of departments around the world, as well as through summer training institutes like IQRM. But, has any of this renewed rigor and systematization of qualitative methods reached our undergraduates? Let me review the results of some of my work documenting our profession’s practice in Scope and Methods courses to help us assess the current state of qualitative methods training for undergraduates. I find that while qualitative methods are taught with some regularity, there is still much work to do to incorporate them into the undergraduate Scope and Methods course. I conclude with some practical suggestions toward that end.

What Do We Know about Scope and Methods Courses?

When I was an assistant professor at Louisiana State University, my colleagues and I were involved in discussions about requiring a Scope and Methods course for our undergraduates. The assistant professors all thought we should absolutely have such a course, while many (though not all) of our senior colleagues were adamantly opposed to such a curricular intervention. Each side in the debate had plenty of anecdotal evidence to support its position, which led fellow assistant professor Robert Hogan and me to design and field a survey of Political Science departments in 2003. We wanted to know how many, and what types of, departments required Scope and Methods courses. The results of this survey were published in an article in *PS: Political Science and Politics* in 2005.¹

As soon as we had compiled the results, we recognized the source of our competing anecdotal claims about the preva-

lence of such courses in the discipline. Not quite half (47%) of Ph.D. granting institutions required an undergraduate methods course. Students in any other type of institution were significantly more likely to have a required course while studying Political Science. Around 61 percent of B.A. granting Political Science departments had a required course, while students in B.A. Social Science (74%) or combined programs (e.g., Political Science and History) (78%) or M.A. Programs in Political Science (79%) were even more likely to be required to take an undergraduate methods course.

We found these results quite perplexing since Ph.D. institutions are the primary generators of knowledge about political science. It made some sense to us that combined, interdisciplinary programs might use methodological training as a unifying force. It also made sense that M.A. programs would require methods in order to prepare their graduates for entrance into Ph.D. programs. It *did not* make sense to us that less than half of Ph.D. institutions required methodological training at the undergraduate level in Political Science.

We hypothesized that this might be a resource issue, but we found that schools with less faculty resources were actually more likely to require a methods course. Smaller programs, such as those offering only a B.A. or M.A., clearly were willing to devote scarce faculty resources to methodology, while over half of the Ph.D. programs were not. This result may help to explain why most of the debate in the literature about appropriate research methodology training has been confined to the graduate level. We concluded our analysis with a plea that Ph.D. institutions consider devoting resources to the teaching of undergraduate methods courses; however, we did not explore the content of such courses in that article.

Where do Qualitative Methods Fit into Scope and Methods courses?

In a follow-up article, Charles Turner and I began to probe what topics are typically covered in undergraduate Scope and Methods courses.² We conducted a second survey of instructors of Scope and Methods courses to see what they taught, what books they used, and even what their courses were typically labeled. The most common titles for these courses included “Research Methods” (41%), “Introduction to Political Research/Analysis” (16%), “Political Analysis/Inquiry” (12%), and “Scope and Methods” (9%). The most popular books used in these courses included *The Essentials of Political Analysis* by Mark Pollock and *Political Science Research Methods* by Janet Buttolph Johnson and H. T. Reynolds, each garnering 16% of the market, though a wide variety of texts and books were reported to be in use by the survey respondents.

The most revealing information came in the description of the topics covered in the course. Nearly all courses (>90%) covered issues related to measurement, the elements of research design, the logic of scientific reasoning, causality, sampling, and survey research. Most (>60%) covered the components of a research paper, experiments, the use of existing data sets, the nature of social science, quasi-experiments, quantitative data analysis, and research ethics. Beyond the basics of social science research, what most of us would consider spe-

cific qualitative methods skills and training is less frequent. The responses included interview techniques (57.1%), case studies (54.3%), comparative method (43.8%), archival research (41%), qualitative data analysis (38.1%), and ethnographic/field research (36.2%). Overall, about three-quarters of the Scope and Methods courses taught quantitative data analysis with the use of statistical software, while only around 40 percent of these courses included instruction in qualitative analysis of one sort or another. The news is therefore not as bad as it could be, nor is it as good as one would hope for the instruction of qualitative methods.

We did not explore the divide between the prevalence of teaching quantitative and qualitative methods in the article. Why might there be such a difference? I can offer some suggestions based on my many years of involvement with the American Political Science Association’s Teaching and Learning Conference (TLC). I participated in, or moderated, the Teaching Research Methods track for several years. Several themes repeatedly emerged across different groups of individuals in this track through the years. First, teaching a Scope and Methods course is challenging in terms of the wide variety of material that could or should be covered. This is exacerbated by the fact that most departments only have one course on Scope and Methods at the undergraduate level. In many departments this course must cover quantitative techniques, which essentially transforms the sole course into research design plus statistics. Attempting to squeeze in a full or even partial repertoire of qualitative methods becomes quite difficult in a one-semester course like this. Second, the instructors for these courses are often those who specialize in quantitative techniques. These instructors thus teach to their strengths and focus primarily on quantitative analysis because they themselves do not have adequate training in qualitative methods. Third, although the respondents to our survey indicated a great deal of consensus about what should be taught in Scope and Methods courses, we are still lacking a disciplinary consensus on best practices for such courses. Recent discussions in the APSA TLC Teaching Research Methods track have called for the incorporation of a variety of quantitative and qualitative methods. Finally, many have argued that methodology should be taught across the curriculum, rather than focused in one course. It is possible that qualitative methods are already being taught in substantive seminars, rather than in a Scope and Methods course.

The Way Forward

Given the flourishing of scholarship on qualitative research methods in recent years, it only makes sense to include systematic training on these topics in undergraduate Scope and Methods courses. The contemporary array of qualitative methods has come a long way in terms of their rigorous application to explaining politics. As more scholars learn about these techniques, we should expect to see them incorporated more fully at the undergraduate level. CQRM and its associated IQRM have had a tremendous impact in moving qualitative scholarship and methodological training forward, but there is more to do.

Announcements

APSA Short Courses Created (or Co-Organized) by Division 46: Qualitative and Multi-Method Research Wednesday, August 29, 2012, New Orleans, Louisiana

Short Course #1: Multi-Method Research

Time: 9:00 am–1:00 pm

Instructors: David Collier, University of California, Berkeley (dcollier@berkeley.edu); Thad Dunning, Yale University; Jason Seawright, Northwestern University; Sheryl Zaks, University of California, Berkeley.

Attention has increasingly focused on how qualitative methods can be linked to other analytic tools, including large-N quantitative analysis and formal modeling. To this end, methodologists have urged scholars to “nest” their case studies within small- to medium-N comparisons, and/or within large-N quantitative analysis. Given that many political scientists are now convinced that good research necessarily employs multiple methodologies, how can different approaches be combined to maximize analytic leverage? How useful are the multi-method techniques under discussion here? Is it sometimes better to stick with one method, and to focus on using it with great skill? This short course explores alternative examples and strategies of multi-method research, with the goal of addressing these questions.

Short Course #2: Designing and Conducting Field Research

Time: 2:00 pm–7:00 pm

Instructors: Diana Kapiszewski, University of California, Irvine, dianakap@uci.edu; Naomi Levy, Santa Clara University, nlevy@scu.edu.

This short course addresses a variety of field methods and data collection techniques, and aims to help scholars hone their empirical research skills. Two foundational premises of the course are that fieldwork “begins” long before one enters the field, and that the way fieldwork is carried out can have important ramifications for data-analysis and theory-generation. We will begin by examining different “varieties” of fieldwork and exploring how research design and fieldwork intersect—since planning for the effective use of field methods and the efficient collection of data are crucial aspects of overall research design. The bulk of the course will focus on preparing for field research, including both logistical and intellectual planning; and conducting field research, including both conversational (interviewing, oral histories, and focus groups) and non-conversational (using archival sources, collecting documents and ephemera, and ethnographic study) forms of data collection. Finally, we will discuss challenges involved in managing, analyzing, and evaluating data both in and out of the field. Although field methods are usually associated with “studying politics abroad,” we discuss techniques that may be applied inside and outside the U.S. The course will include several hands-on activities.

Scholars typically initiate their projects by mapping out their analytic questions, thinking about how they will measure their variables and what evidence they will need to support their claims, and beginning to identify potential sources for the data they hope to collect. Yet even if the research is well planned and adequately funded, obstacles can arise. Key respondents may be unhelpful or unavailable. Valuable archives and other collections of primary materials may be accessible only on a limited basis or may be poorly organized. Data

First, IQRM graduates and allied scholars should make an effort to teach undergraduate Scope and Methods courses in their home department. Our experimentation with the right balance of qualitative and quantitative techniques in such courses can usefully show others how it can be done. Posting undergraduate syllabi in the CQRM archive, along with the many existing graduate syllabi, is an institutionalized way to model such courses.

Second, we can lead the way in identifying disciplinary best practices for the undergraduate Scope and Methods course. While there is undoubtedly debate within the qualitative methods community about the core of such requirements, it is better to have that debate from within, than settle on what quantitative scholars may determine is useful. As always, part of the challenge for qualitative scholars is to educate those of a quantitative orientation about how our research helps to inform and complement their work, as well as generate knowledge that is unobtainable through quantitative techniques.

Finally, we can lead the way in introducing undergraduate methodological training throughout the political science curriculum. Thus, even if progress is slow on including qualitative training in the Scope and Methods course, it will surface elsewhere in the curriculum. Scope and Methods courses may ultimately be the site of the basic of social science research and design with quantitative and qualitative methods taught in other places in the curriculum.

Future undergraduates should benefit from a wide array of methodological training. Such training is useful not only for those preparing for graduate and professional school, but for all kinds of employment, as well as the ability to make sense of individual and societal choices we are confronted with on a daily basis. Rigorous qualitative methods training should be an integral part of the Scope and Methods course, or implemented as part of methods training across the curriculum. This is clearly a matter of faculty choice in structuring the curriculum—Ph.D. institutions in particular need to increase attention to methods training, and all programs need to do a better job of striking the right balance of qualitative and quantitative training. Qualitative scholars can help lead the way in setting best practices for the undergraduate Scope and Methods course, so that our students are as prepared as possible for their post-baccalaureate futures.

Notes

¹ Cameron G. Thies and Robert E. Hogan, “The State of Undergraduate Research Methods Training in Political Science.” *P.S. Political Science and Politics* 38:2 (2005), 293–297.

² Charles C. Turner and Cameron G. Thies, “What We Mean by Scope and Methods: A Survey of Undergraduate Scope and Methods Courses.” *PS: Political Science and Politics* 42:2 (2009), 367–373.

necessary for constructing sampling frames for formal or informal interviewing may simply not exist. Time or money may run out before essential data have been collected.

This short course will help analysts to anticipate and address many of the challenges involved in designing and conducting field research. We discuss strategies that will allow scholars to: (1) convert their research design into a “to get” list; (2) identify and begin to investigate data sources before leaving their home institution; (3) make optimal use of relevant technologies (e-mail, web, cell phones, portable photocopying equipment, scanners, digital cameras, and voice and video recorders); (4) respond to the availability of data not anticipated in the original research design, and to the inaccessibility of data that was originally to be collected; (5) organize and manage the potentially vast quantities of information gathered; (6) establish key contacts and interact constructively with actors of all types in the host community; (7) cope with professionally, politically, and personally uncomfortable situations; (8) make the transition from data collection to data analysis and writing in a timely manner.

Following the end of the formal class at 6 pm, the instructors will hold a “workshop” in which short-course participants will have the opportunity to discuss their own research and the design and conduct of their own fieldwork in a smaller-group setting. We encourage students to stay for this more informal conversation, and to bring along questions about their work.

Participants will be provided with document templates that may be useful when carrying out field research, including sample correspondence. The course is valuable for students planning dissertation projects, for scholars who would like to develop or improve their data collection skills, and for those who teach classes on research methods.

Short Course #3: Qualitative Comparative Analysis and Fuzzy Sets

Time: 2:00 pm–7:00 pm

Instructor: Charles C. Ragin, University of Arizona, cragin@email.arizona.edu.

The analytic challenge of case-oriented research is not simply that the number of cases is small, but that researchers gain useful in-depth knowledge of cases that is difficult to represent using conventional forms (e.g., representations that emphasize the “net effects” of “independent variables”). The researcher is left wondering how to represent knowledge of cases in a way that is meaningful and compact, but which also does not deny case complexity. Set-theoretic methods such as Qualitative Comparative Analysis (QCA), the central focus of this workshop, offer a solution. QCA is fundamentally a case-oriented method that can be applied to small-to-moderate-sized *N*s. It is most useful when researchers have knowledge of each case included in an investigation, there is a relatively small number of such cases (e.g., 10–50), and the investigator seeks to compare cases as configurations. With these methods it is possible to construct representations of cross-case patterns that allow for substantial heterogeneity and diversity. This workshop offers an introduction to the approach and to the use of the software package fsQCA (a free download from www.fsqca.com). Both the crisp (i.e., Boolean) and fuzzy-set versions of the method will be presented.

Fuzzy set analysis is gaining popularity in the social sciences today because of the close connections it enables among verbal theory, substantive knowledge (especially in the assessment of degree of set membership), and the analysis of empirical evidence. Fuzzy sets are especially useful in case-oriented research, where the investigator has a degree of familiarity with the cases included in the investigation and seeks to understand cases configurationally—as specific combina-

tions of aspects or elements. Using fuzzy-set methods, case outcomes can be examined in ways that allow for causal complexity, where different combinations of causally relevant conditions combine to generate the outcome in question. Also, with fuzzy-set methods it is possible to evaluate arguments that causal conditions are necessary or sufficient. Analyses of this type are outside the scope of conventional analytic methods.

Specific topics addressed in the course include the differences between set-theoretic and correlational methods; conventional crisp sets versus fuzzy sets; the calibration of fuzzy sets; how calibration differs from conventional forms of measurement; analyzing fuzzy set relations; the correspondence between concepts and fuzzy set membership scores; the correspondence between theoretical statements and the analysis of fuzzy set relations; using fuzzy sets to study cases as configurations; and using fuzzy sets to unravel causal complexity, with a special focus on the equifinality.

APSA Panels/Roundtables Created (or Co-Organized) by Division 46: Qualitative and Multi-Method Research August 30–September 2, 2012, New Orleans, Louisiana

Quantitative Qualitative Methods

Patrick Mello, “A Critical Review of Applications in QCA and Fuzzy-Set Analysis and a ‘Toolbox’ of Proven Solutions to Frequently Encountered Problems.”

Nicholas Knowlton, “Qualitative Comparative Analysis as a Nested Approach to Inquiry.”

James Honaker, “A Quicksort Implementation for Qualitative Data for Recovering Latent Dimensions with Crowdsourced Human Computation: Measuring State Policies for Welfare Eligibility Under TANF.”

Markus Haverland, “Two or Three Approaches to Explanatory Case Study Research?”

The Contributions of Qualitative Methods to Latin American Scholarship

Chair: David Collier

Santiago Anria and Sara Niedzwiecki, “Social Movements and Universal Social Policies: The Bolivian Renta Dignidad.”

Carlos Freytes Frey, “Using Mechanisms to Unpack Concepts: Party Institutionalization and Its Impact on the Latin American Left’s Policy Choices.”

Maria Paula Saffon Sanin, “Theft or Inequality? Identifying the Causes of Land Conflict in Twentieth Century Latin America.”

Maria Agustina Graudy, “A Conceptual Typology of State Territorial Reach: Evidence from Latin America.”

Jennifer Marie Cyr, “Measuring ‘the Social’: The Use of Focus Groups in Social Science Research.”

Discussant: James Mahoney

Conceptualizing State Strength: A Cross-Regional Perspective

Chair: Catherine Boone

Scott Radnitz, “Revolutions and State Weakness.”

Hillel David Soifer, “Conceptualizing and Measuring State Weakness.”

Catherine Boone, “Geopolitical Scale and Politics in Africa: Elections, War, and Public Opinion.”

Juan Pablo Luna, “State Challenges in Contemporary Latin America: A Typology.”

Discussant: Richard Snyder

Combining Large, Medium, and Small-N

- Chair and Discussant: James Mahoney
Sebastian Guzman, "Strikes and Leadership's Ideology: A Critique of Rational Choice Theories of Union-Government Interactions with Evidence from Chile."
Brian Paulmer-Rubin and Anne Meng, "Using In-Depth Case Studies to Evaluate the Plausibility of Ecological Inference in Large-N Research"
Torben Heinze, "Leaders, Followers and Laggards in West European Politics: On Causal Configurations in Education Policy."
Derek Beach, "'And Never the Twain Shall Meet?' Can Process Tracing and QCA Methods be Combined? Lessons from a Case Study of Representation in the EU."

Types, Typologies, and Causal Explanations

- Chair and Discussant: David Waldner
Svend-Erik Skaaning and Jorgen Maller, "Assessing the Explanatory Power of Typological Theories."
Amy Liu and Richard Doner, "Crisis, Shocks, and Threats: A Conceptual Typology of 'Exogenous' Events."
Ruchan Kaya, "A New Type of Resource Curse? Effect of Natural Resource Pipelines on Economic and Human Development."

Process Tracing Corruption and the Provision of Public Goods

- Chair and Discussant: Gerard Alexander
Jaemin Shim, "Process Tracing for Causal Explanation and Causal Inference: A Comparative Case Study of Welfare Legislation Patterns in South Korea and Japan."
Rodolfo Sarsfield, "Is Corruption a Path Dependence Process? Focal Points and Corrupt Behavior."
Carolyn Warner, Ramazan Kilinc, and Christopher Hale, "Intrinsic Religiosity and Public Goods Provision: Evidence from Catholicism and Islam."

Innovations in Qualitative Data Collection and Analysis

- Brendan O'Rourke and John Hogan, "Neo-Liberal Discourse on an Edge: Communicating Political Economic Crisis in Post-2007 Ireland."
Alexander Dunning Jakle, "The Surveyor, the Reporter, and the Detective: Three Models of the Qualitative Interview."
Markus Kreuzer, "Are All Hypotheses Created Equally? Using Footnotes to Replicate Theory Construction."
Shivaji Kumar, "Sequential, Multimethod Identity Research in Security Studies: The Promise of Combining Discourse Analysis and Survey Methods."

Causal Explanations and Causal Mechanisms

- Chair and Discussant: David Waldner
Joseph O'Mahoney, "Why Did You Do That? The Methodology of Motive Attribution."
James Robinson, "Paradox of Confirmation: Causation, Explanation, and Inference in IR."
Miles Townes, "Are Causal Mechanisms Real?"

Roundtable: Field Research and Representing African Politics in the 21st Century

- Chairs: Lauren Maclean and Parakh Hoon
Participants: Catherine Boone; Leonard Wantchekon; John Harbeson; Robert Bates; Lauren MacLean; Parakh Hoon.

Social Science Methods and the Real World

- Chair and Discussant: Evan Lieberman
Ian Lustick, "Validating and Verifying Validation and Verification: The Methodological Challenge of a Public Policy Imperative."
Matthew Tubin, "Dr. Strange-Economist or How I Learned to Stop Worrying and Love Financial Models."
Colin Elman, "Within Case Analysis: How Non-Comparable Observations Contribute to Descriptive and Causal Inference."

Dimensions of Time: Timing, Sequencing, Age, Tempo, and Period

- Chair: Marcus Kreuzer
Andre Kaiser, "When Time is on My Side: Time Horizons and Institutional Reforms"
Marcus Kreuzer, "Time as Age: On the Importance of Duration in Studying Politics."
Tim Luecke, "Staggering and Democratic Timescapes."
Klaus Goetz, "Biographical Questionnaires: How to Seize Numerous Life Histories."
Discussant: John Oates

Strategies of Case Selection

- Chair and Discussant: Jason Seawright
Derek Beach and Rasmus Pedersen, "Case Selection Strategies in the Three Variants of Process Tracing Methods."
Giovanni Capoccia and Laura Stoker, "Choosing and Combining Units in Political Science Research."
Gary Goertz, "Case Studies, Causal Mechanisms, and Selecting Cases: Most Similar Systems, Control Variables, and Matching."
Eric Grynaviski, "Selecting Cases that Travel: A Distributed Cognition Approach to Case-Study Research."

Critical Concepts: Authoritarianism, Corruption, and Representation

(Co-sponsored with IPSA Research Committee 1 [Concepts and Methods])

- Chair: Staffan Lindberg
Daniel Neep, "An Interpretive Approach to Concept Formation in the Study of Authoritarianism: Abstraction, Historicism, and Ideal Types in Political Science."
Yonatan Morse, "Conceptual Advances in the Study of Electoral Authoritarianism—Disaggregating Competitiveness and Hegemony."
William English, "Varieties of Corruption and the Architecture of Public Trust."
Sayres Rudy, "Glitches in the Matrix: Suturing the Conceptual Divide toward Explanations of Re-representation."
Discussant: Hans-Joachim Lauth

The Methods Café

Emily Hauptmann, “Archival Research.”
Ronald Schmidt Sr., “Critical-Interpretive Policy Analysis.”
Mary Hawkesworth, “Feminist Methods.”
Dorian Warren and Katherine Walsh, “Field Research.”
Peregrine Schwartz-Shea and Cyrus Zirakzadeh, “Interpretive Research Design.”
Ange-Marie Hancock, “Intersectionality Research.”
Robert Kaufman Adcock, “Why Compare? Comparison and Concepts in Interpretive Research.”

Intersectionality: Interpretations and Interpretive Methods/Methodologies

(co-sponsored by Related Group on Interpretive Methodologies and Method and Section on Women and Politics)

Chairs: Julia S. Jordan-Zachery and Andrea Y. Simpson
Amy Cabrera Rasmussen, “Interpretive Category Analysis: Insights and Implications for Intersectionality Research.”
Anna Sampaio, “Intersectionality as Methodology.”
Edwina Barvosa, “Interpreting Intrapsychic Intersectionality: Reading Queer Desire in Anzaldúa and Shakespeare’s Sonnet 20.”
Discussants: Nikol G. Alexander-Floyd and Rita Dhamoon

Contributions of Interpretive Methods to Political Analysis

(co-sponsored by Related Group on Interpretive Methodologies and Method and IPSA Research Committee 1 [Concepts and Methods])

Chair: Charles L. Mitchell
David V. Edwards “Using Interpretive Methods to Create More Useful Policy Theory.”
Cyrus Ernesto Zirakzadeh, “Placing Political Pronouncements in Their Cultural Contexts: A Comparison of Ricoeur’s Theory of Narrative and Zizek’s Vision of Habit.”
Ragnhild Louise Muriaas, “Advocating for Comparative Interpretive Analysis (CIA): Lessons From Studying Local Elections in Malawi, South Africa, and Uganda.”
Elena Gadjanova, “I Say a Word, They Hear a Story: Understanding how Political Communication Evoking Individual Identities Constructs Political Communities.”
Discussant: Hans-Joachim Lauth

**Berlin Summer School in Social Sciences:
Linking Theory and Empirical Research
Berlin, July 15–27, 2012**

The Berlin Summer School in Social Sciences: Linking Theory and Empirical Research aims at promoting young researchers by strengthening their methodological understanding in linking theory and empirical research.

In a first step, we tackle the key methodological challenges of concept-building, causation, and micro-macro-linkage that occur in all research efforts and aim at a clarification of the epistemological implications underlying methodological paradigms. In a second step, we apply these methodological considerations by looking at how central empirical fields of research in political science and sociology have dealt with these challenges and—by referring to selected empirical studies—what solutions have been found to overcome them. Furthermore, participants are provided with hands-on research advice and have the opportunity to present their own work and approaches to these issues.

The Berlin Summer School is a joint endeavor of two of Germany’s leading social science institutions, the Berlin Graduate School of Social Sciences (BGSS) at Humboldt-Universität zu Berlin and the Social Science Research Center Berlin (WZB).

The two-week summer school both attracts internationally renowned scholars and draws on Berlin-based faculty. Among the confirmed international lecturers are Delia Baldassarri (Princeton University), Peter Bearman (Columbia University), Mark Bevir (University of California, Berkeley), Henry E. Brady (University of California, Berkeley), Craig Calhoun (New York University/London School of Economics), Donatella Della Porta (European University Institute), Ronald Inglehart (University of Michigan), and Klaus von Beyme (Universität Heidelberg).

For additional information, please visit our webpage at www.berlinsummer-school.de or contact directly Andreas Schäfer at summer-school.bgss@hu-berlin.de.

Letter from the Editor continued from page 1

paragraph to illustrate contemporary trends, but I am sure the list is not exhaustive, and I encourage any of you with forthcoming method and/or methodology books anywhere within the epistemically pluralistic reach of our section, do let me know about them!

Intellectual change comes not only from new publications, but also new initiatives and changing norms. The second symposium in this issue informs readers about a major new conversation on qualitative data access, being undertaken with support from the NSF and the APSA. While part of our membership may already be aware of this, it is something that should be on the radar screen of everyone whose research involves qualitative methods, in any of their varied forms. Finally, this issue also carries forward the attention to pedagogical matters, which has been a recurring mission of the newsletter from its earliest days, with an agenda-setting discussion by Cameron Thies of methods teaching at the undergraduate level. As usual with the spring issue, the newsletter includes an overview of the panels the section, and selected related groups, will sponsor at the annual APSA meeting, which this year will be in New Orleans.

As the recently installed third editor of this newsletter, I would like to close my first Letter from the Editor by thanking, on behalf of the section, my two predecessors—John Gerring and Gary Goertz—for their labors in making this publication so consistently excellent. As their successor, I aim to carry forward the newsletter's best traditions of presenting substantively rich and diverse material in a lively, at times even opinionated, first-person style, free from certain constraints of journal prose. The key to ongoing vitality of the newsletter is ultimately, however, engagement from our section membership. Via your section dues (thank you!) every one of you helps make this newsletter possible. I strongly encourage you not only to read the newsletter you fund, but also to think about it as a venue that you might shape and contribute to. If you have ideas about symposia that you would like to propose and organize, other potential articles, announcements, etc., do not hesitate to e-mail me.

Qualitative and Multi-Method Research

The George Washington University
Department of Political Science
Monroe Hall 440
2115 G Street, N.W.
Washington DC 20052
USA

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

Qualitative and Multi-Method Research (ISSN 2153-6767) is edited by Robert K. Adcock (tel: 510-301-4512, fax: 202-994-7743, email: adcockr@gwu.edu). The production editor is Joshua C. Yesnowitz (email: jcyesnow@bu.edu). Published with financial assistance from the Consortium for Qualitative Research Methods (CQRM). Opinions do not represent the official position of CQRM. After a one-year lag, past issues will be available to the general public online, free of charge, at <http://www.maxwell.syr.edu/moynihan/programs/cqrm/section.html>. Annual section dues are \$8.00. You may join the section online (<http://www.apsanet.org>) or by phone (202-483-2512). Changes of address take place automatically when members change their addresses with APSA. Please do not send change-of-address information to the newsletter.