

Data Access and Research Transparency in the Qualitative Tradition

Colin Elman, *Syracuse University*

Diana Kapiszewski, *Georgetown University*

As an abstract idea, openness is difficult to oppose. Social scientists from every research tradition agree that scholars cannot just assert their conclusions, but must also share their evidentiary basis and explain how they were reached. Yet practice has not always followed this principle. Most forms of qualitative empirical inquiry have taken a minimalist approach to openness, providing only limited information about the research process, and little or no access to the data underpinning findings. What scholars do when conducting research, how they generate data, and how they make interpretations or draw inferences on the basis of those data, are rarely addressed at length in their published research. Even in book-length monographs which have an extended preface and footnotes, it can sometimes take considerable detective work to piece together a picture of how authors arrived at their conclusions.

There are multiple overlapping reasons why scholars might follow this minimalist approach. One root is a “craft” understanding of qualitative research. According to this view, only scholars who are as immersed in the background and detail of an author’s cases, and as familiar with the bulk of the evidence as he or she is, are in a position to make an informed judgment about the conclusions. Given that those scholars are well-equipped to locate the relevant sources and understand their connection to claims, brief footnotes are deemed sufficient. In addition, because the discipline does not currently reward greater openness, scholars are reluctant to take on the logistical burden it seems to imply in the absence of a clear and substantial payoff. Other objections arise because qualitative data are often under constraint. For example, concerns about human subjects may limit sharing when it would impinge on the rights and welfare of study subjects. Copyright and other proprietary rights might also limit transparency as many sources may not be in the public domain.

We are not wholly convinced by these justifications for taking a minimalist approach to openness. We doubt that the level of detail most qualitative researchers currently provide about their data, and about the connections between their data and their conclusions, is sufficient for even well-informed interlocutors to fully appreciate their arguments. Given the transaction costs of “looking for themselves,” even specialists end up relying on a trust norm and heuristic shortcuts.¹ Further, as the discipline becomes more aware of the benefits of openness, we anticipate that gatekeepers such as journals, funding

agencies, and academic departments will increasingly implement openness as a merit criterion—meaning that for individual scholars, adopting transparency practices sooner rather than later is actually the rational route to take. Finally, while we recognize that human subjects and copyright concerns may place limitations on how much openness can occur, we believe that a great deal of information can be shared without crossing either ethical or legal boundaries.²

Starting from the position that qualitative research as it is currently practiced is a valuable enterprise, in this article, we investigate whether and how qualitative scholars can reveal more about the processes through which they generate, analyze, and deploy data. Our central message is that if qualitative scholars take a more self-conscious, deliberate, and expansive approach to data access and research transparency (DA-RT), they can demonstrate the power and rigor of their work more clearly and empower a much larger audience to understand and interpret their research on its own terms.

OPENNESS AS A “META STANDARD”

All rule-based social inquiry is based on three notions. First, scholarly communities hold shared and stable beliefs that research designed and conducted in particular ways—according to particular rules—is warranted to produce knowledge with certain characteristics. Second, both the conduct of social inquiry and the written products that represent its conclusions are designed to capture those characteristics. Finally, to possess those characteristics, research must be designed and conducted in accordance with those rules.

Thus when scholars claim to have explained, interpreted, predicted, or otherwise asserted knowledge about an aspect of the social world, the warrant for making that claim does not come solely from the data they collected and analyzed. It comes in part from theories of knowledge that argue that when data generation and analysis follow certain rules, such claims can be made. In turn, our evaluation of those claims is based on our assessment of whether scholars’ research processes followed those rules.

Consider, for example, the connection between epistemology, method, and results in experimental research (Druckman, Green, Kuklinski, and Lupia 2011, 17). The departure point for such research is that randomized assignment addresses the fundamental problem of causal inference and that any difference in outcome between a control and treatment group will be, on

average, attributable to the treatment, that is, to the causal intervention. Hence research designs can be meaningfully divided into those that can establish causation and allow researchers to aver that they have done so (if the experiment is designed and conducted according to the relevant rules), and those that cannot.

Openness allows authors an opportunity to show they are conducting inquiry of a certain type, and thus can potentially make claims with particular characteristics. Derivatively, openness allows scholars to show that a particular piece of research was well designed and done correctly, and hence commands the virtues of those additional warrants. That is, our judgments as to the relative strength of research amount to assessments of whether it was produced in accordance with the relevant rules.

Foundational epistemologies, and the characteristics of their archetypal knowledge statements, differ across research communities. Nevertheless, as Lupia and Elman (2014, this symposium) note, at a general level, some common ground exists among them. The methodologies political scientists use to reach evidence-based conclusions *all* involve extracting

information from the social world, analyzing the resulting data, and reaching a conclusion based on that combination of the evidence and its analysis.

approach to qualitative inquiry, process tracing. We use this example to illustrate the underlying connections among epistemology, analytic method, data, and conclusions. Scholars who use process tracing engage with the social world, draw observations from it, generate data, conduct analyses, and deploy data to support their claims in particular ways. These contrast both with how scholars conducting large-N observational research carry out these tasks and how scholars operating in alternative sub-types of qualitative inquiry do so.⁴ Given their varying foundations, these other approaches, by definition, take a different view of data and analysis. Correspondingly, their discussions of transparency would be different.

To be clear, our intention is not to provide a full account of process tracing itself. Rather, we hope to illustrate how greater transparency would make the technique's strengths more apparent, and focus critics on its real rather than its imaginary weaknesses.

To scholars unfamiliar with qualitative research, process tracing might seem like unsystematic soaking and poking

When scholars claim to have explained, interpreted, predicted, or otherwise asserted knowledge about an aspect of the social world, the warrant for making that claim does not come solely from the data they collected and analyzed. It comes in part from theories of knowledge that argue that when data generation and analysis follow certain rules, such claims can be made.

Hence, and as reflected in the revisions to APSA's ethics guide (Lupia and Elman, *ibid.*), openness requires all scholars to provide access to the data on which their conclusions were based and to clearly describe the analysis that produced those conclusions. Yet despite the universal applicability of openness as a meta standard, transparency is always instantiated locally.³ That is, openness requires everyone to show their work, but what they show and how they show it varies. These differences are grounded in epistemic commitments and the rule-bound expectations of the tradition in which scholars operate.

for evidence that substantiates an empirical claim. In part, this misimpression is a function of process tracing being more sophisticated than its typical representation in published substantive applications suggests. As occurs with much qualitative work, the rigorous techniques that underlie process tracing often remain implicit in published scholarship (Mahoney 2012, 14, 21), rendering it vulnerable to critique. Transparency calls on qualitative scholars to make these techniques more explicit.

Among qualitative methodologists there is a widespread consensus that single pieces of data play a crucial role in process tracing.⁵ Our key point is this: identifying those "diagnostic" data is an *analytic* procedure that scholars who use the technique should clearly describe in their work.

A qualitative datum, like a quantitative datum, is one among many: both are considered in the context of many other pieces of information. How the single piece of information relates to the group, however, is quite different in the two modes of analysis. A quantitative datum is one among many of the same thing, a comparable measure of the same characteristic in a sample. In process tracing, a qualitative datum is a single, unique piece of information that nevertheless gathers its meaning as part of the larger constellation of data in which it is embedded.

To be sure, as Andrew Moravcsik (2014, this symposium) persuasively argues, scholars should show the data they cite, and make clearer how those data support their claims.

RESEARCH TRANSPARENCY AND PROCESS TRACING: AN EXAMPLE

We have argued that how openness is achieved in a particular tradition depends on how its scholars generate and use data to gain inferential or interpretive leverage. Developing transparency practices for qualitative research thus requires understanding how (and why) the observations scholars draw from the social world are converted into data, how a subset of those data are used in support of their analytic claims, and how a subset of *those* data are selected for citation.

In this section, we discuss some issues that arise when pursuing transparency in the context of one within-case

Scholars should also, however, give some consideration to representing more of the body of material they consulted during research. Indeed, because qualitative scholars engage with the social world (rather than with a discrete data set), they inevitably encounter many more potentially relevant sources than they engage, draw many more observations from those sources than they convert into data, and generate more data than they cite in their research.

A cited datum has a complicated relationship with that larger uncited corpus. When scholars assert the importance of a qualitative datum drawn from a particular source by citing it, they are almost always implicitly making two statements about the *other* sources they consulted: observations drawn from those other sources contributed to giving the cited source its meaning, *and* none of these other sources contains more diagnostic information than the one cited. Thus what appears to be a wholly granular use of data in a process

To be clear, we are not suggesting that process tracing—or any other qualitative analytic technique—be done any differently than it is now. We simply encourage scholars to be clearer about what they did and to share more of the data that underlie their claims. In other words, we hope to expose the aspects of qualitative work that often remain invisible, thus helping qualitative scholars to convey the power of their research.

EVALUATING QUALITATIVE RESEARCH

Transparency is also a prerequisite for meaningful evaluation. As we noted earlier, claims and inferences produced by social inquiry are only valid to the extent that the work that produced them followed the rules of the relevant research tradition. Evaluating research thus entails assessing whether it was designed and conducted in ways that followed those rules.

In contrast to some other research traditions, qualitative scholars are only partially committed to replication as a mech-

Among qualitative methodologists there is a widespread consensus that single pieces of data play a crucial role in process tracing. Our key point is this: identifying those “diagnostic” data is an analytic procedure that scholars who use the technique should clearly describe in their work.

tracing study is actually somewhat holistic. Moreover, how a datum comes to be considered diagnostic is explicitly a product of research design. Hence, scholars writing process tracing accounts triage data to identify clearly delineated tests of necessity and sufficiency. It is because some observations are usable in hoop or smoking-gun tests that they are probative for the hypothesized explanations. That is, the logical requirements of the research design identify which data are diagnostic.⁶

Given the relevance of the large body of material consulted during research to a process tracing study, its partial representation in most such accounts runs counter to the nature and spirit of qualitative inquiry. It leads to an understatement of the contribution made by the broad set of data that qualitative scholars generate. It also compromises the persuasiveness of qualitative accounts. Traditionally cited data cannot fully represent the breadth of material used to draw inferences and arrive at conclusions in qualitative research. Of course it is impossible for scholars to show all of the data they used in an analysis; however, a project's evidentiary base—and the relationship between cited data and the broader set of data used in the analysis—can be more carefully described or at least better referenced.⁷

As APSA's ethics standards suggest, achieving transparency in a published process tracing account thus requires scholars do more on three fronts. They should more explicitly describe how they drew observations and generated data,⁸ more precisely explain how they deployed those data and used process tracing to reach their conclusions, and share more of those data, than is currently the norm.

anism for establishing validity. This is especially true for evaluating data *generation*. Rarely does one qualitative scholar “redo” another's interviews, for instance—because sufficient information is not generally provided to do so and because the logistical burden is often prohibitively large. In addition, it would be unrealistic to expect the exact same information to be garnered given how contextual variables affect such interactions.⁹ But, even without a commitment to replication, openness is necessary for readers to assess how scholars drew observations from sources, attached meaning to them, and identified them as analytically significant to their research.

For instance, readers can carefully assess whether the authors' data-generation techniques were aligned with the rules of inference and interpretation they were following. This alignment is a predicate requirement for the associated analytic methods to be successfully employed. Readers will also want to assess whether any given data-generation technique was used effectively. Hence, they might see what they can learn about how authors drew observations from sources, for instance, reading their interview transcripts to assess whether they asked leading questions or in some other way biased their interviews. Readers could also assess whether researchers engaged in triangulation, given the subjectivity inherent in many qualitative sources. At an extreme, readers could go to the research context in which authors worked and draw observations from different sources (e.g., interview a different set of relevant actors) and evaluate the consistency between those observations and those drawn from cited sources (i.e., engage in post hoc triangulation). As DA-RT practices develop, standards and norms for systematically generating data in qualitative work will become more explicit,

and scholars' data-generation techniques will be more easily evaluated against them.

With regard to evaluating the data *analysis* underlying the conclusions drawn in qualitative work, the most plausible standard is whether a reader could analyze the data cited as evidence and arrive at the same conclusions. Evaluating data analysis might entail assessing whether the operations authors performed on their data were appropriate (i.e., were dictated by the rules undergirding the form of analysis in which they were engaging). It might also involve assessing the micro-connections between individual pieces of (cited) data and descriptive/causal inferences or interpretation, again in view of the analytical methods the scholars were using, to determine whether the data support the inferences. Returning to the process tracing example, have the data cited been used in rigorous tests that confirm or disconfirm the competing hypotheses? As this discussion implies and as we suggested earlier, data sharing and analytic transparency are prerequisites for evaluating data analysis.

CONCLUSION: THE VALUE OF OPENNESS

Enhancing DA-RT has the potential to yield significant benefits for qualitative scholars and scholarship. DA-RT empowers researchers to provide a more complete description of the value-added their immersion in the social world provides. Although the details differ across research traditions, DA-RT

DA-RT allows qualitative scholars to demonstrate the power of their inquiry, offering an opportunity to address a central paradox: that scholars who value close engagement with the social world and generate rich, thick data rarely discuss the contours of that engagement, detail how they generated and deployed those data, or share the valuable fruits of their rigorous labor.

allows qualitative scholars to demonstrate the power of their inquiry, offering an opportunity to address a central paradox: that scholars who value close engagement with the social world and generate rich, thick data rarely discuss the contours of that engagement, detail how they generated and deployed those data, or share the valuable fruits of their rigorous labor.

This article focused on transparency techniques in the context of one within-case approach to qualitative research: process tracing. Other traditions have different understandings of how to be transparent about their research practices and grapple with their own difficult questions. Given these different concerns, the transparency conversation will be more productive if scholars from diverse research communities participate and begin to identify the levels and types of transparency with which they are comfortable and that are consistent with their modes of analysis. Thus the ideas we have proposed in this article represent one contribution to a broader ongoing debate that we hope will result in the gradual introduction and acceptance of transparency as a worthy goal, and useful standard, for social science research.

ACKNOWLEDGMENTS

The authors thank Arthur Lupia, Alan Jacobs, and the other participants at the Institute for Qualitative and Multi-Method Research authors' workshop (Syracuse University, June 2013) as well as those who participated in and attended the "Data Access and Research Transparency (DA-RT) in Political Science" committee panel at the APSA Annual Meeting (Chicago, August–September 2013), for helpful comments on earlier drafts. Our point of departure for this article is the recent APSA dialogue on DA-RT, and the article's arguments are partly drawn from an ongoing project on research transparency in the qualitative tradition. Because both projects are mid-course, our arguments should be regarded as works-in-progress as well. ■

NOTES

1. For example, they might use extra-textual clues such as the number of interviews an author conducted or the time she spent in an archive to help assess whether she approached her task in a way that would have allowed her to generate useful data, effectively interpret them, and draw valid inferences.
2. For instance, soliciting appropriate informed consent (referencing data sharing) from project participants, effectively using anonymization strategies, controlling access to data, using precise data-use agreements, and sharing data in line with fair use standards can mitigate the concerns we highlighted previously. The bottom line, of course, is that scholars should only make qualitative data available in ways that conform to ethical and legal imperatives.
3. Sports and law provide analogues: all sports are governed by a meta standard of fair play, and all jurisdictions are governed by a meta standard of legality—yet what this means in practice differs from sport to sport, and jurisdiction to jurisdiction, because the rules differ.
4. Others include counterfactual analysis (e.g., Sekhon 2004), Qualitative Comparative Analysis and other forms of analysis with set-theoretic foundations (e.g., Goertz and Mahoney 2012; Grofman and Schneider 2009; Ragin 2000, 2008; Schneider and Wagemann 2010), and ethnographic analysis (e.g., Schatz 2009; Yanow and Schwartz-Shea 2006).
5. See Beach and Pedersen 2013; Bennett and Checkel forthcoming; Bennett 2010; Collier 2011; Mahoney 2012.
6. This is analogous to the increasing prevalence among quantitative scholars of the view that strong causal inference requires either data produced by an experiment or observational data from circumstances that mimic an experiment (i.e., a natural experiment).
7. Given the increasing use of web appendices and data archiving, the initial clause of this sentence may soon be inaccurate. Data and data collection practices can be described in the "overview" portion of a Transparency Appendix accompanying the published work of scholars engaging in active citation; see Moravcsik (2014, this symposium).
8. Indeed, qualitative scholars' penchant for generating their own data makes clearly describing how, when, where, and why data were generated—the "production transparency" aspect of DA-RT—particularly critical for research transparency in qualitative work.
9. Other forms of data collection may be more easily replicable. It is an open question whether archival research is more replicable than interviews. Assuming universal archival coverage is not possible, different researchers

are likely to take distinct trajectories through the materials and arrive at dissimilar interpretations of them.

REFERENCES

- American Political Science Association. 2012. *A Guide to Professional Ethics in Political Science*, Second Edition. Washington, DC: American Political Science Association.
- Beach, Derek, and Rasmus Brun Pedersen. 2013. *Process-Tracing Methods: Foundations and Guidelines*. Ann Arbor: University of Michigan Press.
- Bennett, Andrew. 2010. "Process Tracing and Causal Inference." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Second Edition, eds. Henry E. Brady and David Collier. Lanham, MD: Rowman and Littlefield.
- Bennett, Andrew, and Jeffrey Checkel, eds. Forthcoming. *Process Tracing in the Social Sciences*. New York: Cambridge University Press.
- Collier, David. 2011. "Understanding Process Tracing." *PS: Political Science and Politics* 44 (4): 823–30.
- Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia. 2011. "Experiments: An Introduction to Core Concepts in Political Science." In *Cambridge Handbook of Experimental Political Science*, eds. James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia. New York: Cambridge University Press.
- Goertz, Gary, and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.
- Grofman, Bernie, and Carsten Schneider. 2009. "An Introduction to Crisp Set QCA, with a Comparison to Binary Logistic Regression." *Political Research Quarterly* 62 (4): 662–72.
- Lupia, Arthur, and Colin Elman. 2014. "Openness in Political Science: Data Access and Research Transparency." *PS: Political Science and Politics* 47 (1): this issue.
- Mahoney, James. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods & Research* 20 (10): 1–28.
- Moravcsik, Andrew. 2014. "Transparency: The Revolution in Qualitative Research." *PS: Political Science and Politics* 47 (1): this issue.
- Ragin, Charles. 2000. *Fuzzy Set Social Science*. Chicago: University of Chicago Press.
- . 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.
- Schatz, Edward, ed. 2009. *Political Ethnography: What Immersion Contributes to the Study of Power*. Chicago: University of Chicago Press.
- Schneider, Carsten, and Cladius Wagemann. 2010. "Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets." *Comparative Sociology* 9 (3): 397–418.
- Sekhon, Jasjeet S. 2004. "Quality Meets Quantity: Case Studies, Conditional Probability, and Counterfactuals." *Perspectives on Politics* 2 (2): 281–93.
- Yanow, Dvora, and Peregrine Schwartz-Shea. 2006. *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*. Armonk, NY: ME Sharpe.