Title
Problematic Tools: Introduction to Symposium on Set Theory in Social Science

Journal
Qualitative & Multi-method Research, 12(1)

Author
Collier, David

Publication Date
2014-07-01

Peer reviewed

Qualitative & Multi-Method Research

Spring 2014, Vol. 12, No. 1

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

Symposium. The Set-Theoretic Comparative Method: Critical Assessment and the Search for Alternatives

Problematic Tools: Introduction to Symposium on Set Theory in Social Science

David Collier
University of California, Berkeley
dcollier@berkeley.edu

> "To welcome ideas that shift the grounds on which our arguments previously found traction—that is our obligation as intellectuals." Wedeen (2014: 1)

Analysts who developed the set-theoretic comparative method (STCM) have formulated admirable goals for researchers who work in the qualitative and multi-method tradition. This method includes above all Charles Ragin's innovative approach of Qualitative Comparative Analysis (QCA), along with further systematization of the set-theoretic framework by other authors.[1] These colleagues are outstanding scholars and intellectual leaders in the field of methodology, and their advocacy of these goals is a major contribution.

However, the analytic tools employed by STCM have in many ways become an obstacle to achieving these admirable goals. For example, the system of fuzzy-set scoring appears to be problematic, poorly matched to a standard understanding of conceptual structure, and perhaps unnecessary in its present form. Computer simulations suggest that findings suffer from serious problems of stability and validity; and while the choice of simulations that match the method is a matter of some controversy, the cumulative weight of simulation results raises major concerns about STCM's algorithms—i.e., its basic, formalized analytic procedures.

Questions also arise about the cumbersome formulation of findings in what is often a remarkably large number of causal paths. Relatedly, some scholars question the STCM's rejection of the parsimonious findings, in the form of "net effects," routinely reported in other methodological traditions. Regarding applications, readily available software has encouraged publication of dozens of articles that appear to abandon key foundations of the method and rely far too heavily on these algorithms. Finally, STCM appears inattentive to the major, recent rethinking[2] of standards and procedures for causal inference[3] from observational data.

These problems raise the concern that the set-theoretic comparative method, as applied and practiced, has become disconnected from the underlying analytic goals that motivated Charles Ragin to create it.

This symposium explores these problems and seeks to identify promising directions for further work that pursues these same goals. In the symposium, this overall set of methods is referred to as STCM, and the designation QCA is used when the discussion is specifically focused on Ragin's contribution. For the convenience of readers, in anticipation that this essay might be read apart from the symposium, full citations to the other contributions are included in the bibliography.

Readers familiar with *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Brady and Collier 2004, 2010) will recognize the parallel with the present symposium. *Rethinking Social Inquiry* addressed an earlier, constructive initiative to redirect thinking about qualitative methods: King, Keohane, and Verba's (1994) *Designing Social Inquiry*—widely known as KKV. Their book had excellent overall goals, which centrally included a concern with systematizing qualitative research procedures that too often are unsystematic and unstandardized.[4] However, the book advocated specific tools for pursuing these goals that many scholars considered inappropriate, and in some respects counter-productive. *Rethinking Social Inquiry* sought to formulate methodological priorities and analytic tools more appropriate to qualitative research.

This symposium adopts the same perspective on the set-theoretic comparative method. The overall goals are excellent, and they centrally include a concern with systematizing qualitative research procedures that too often are unsystematic and

[1] Ragin 1987, 2000, 2008; and above all Goertz and Mahoney 2012, and Schneider and Wagemann 2012. QCA is understood here to include the crisp-set, multi-value, and fuzzy-set versions—i.e., csQCA, mvQCA, and fsQCA.

[2] This rethinking is discussed in Tanner's (2014) contribution to this symposium and in Collier (2014).

[3] The term causal inference is employed by some STCM authors (e.g. Goertz and Mahoney 2012; Schneider and Wagemann 2012), yet for other authors "causal interpretation" and "causal recipe" are preferred. The present discussion respects these distinctions, and uses "causal inference" as an umbrella term that encompasses these alternatives.

[4] David Laitin (1995), well known as a (creatively) eclectic scholar who is deeply engaged in both the qualitative and quantitative traditions, praised KKV as an important step toward "disciplining" political science.

unstandardized. However, the specific tools advocated for pursuing these goals have again been seen by many scholars as inappropriate, and in some respects counter-productive. Finally, in parallel, the symposium explores alternative tools that hold promise for more effectively pursuing these same goals.

The contributors to this symposium hope their essays will move the discussion forward, thereby seeking to sustain the same constructive spirit that the Brady and Collier volume sought to achieve. For both debates, a central recommendation is a return to more traditional qualitative methods,[5] which have in fact seen valuable innovation in recent years.[6]

Excellent Goals

There should be wide agreement that the set-theoretic comparative method has productively extended the horizon of scholars concerned with qualitative and multi-method research. STCM has introduced important new insights and challenged scholars to think about them carefully.

One example is the focus on asymmetric causation,[7] readily understood in terms of two types of causes: blocking causes that prevent a given outcome—as would occur with the absence of a necessary condition; versus triggering causes that ensure its occurrence—as would occur with the presence of a sufficient condition. The importance of this idea is seen in the fact that for many political scientists, it initially produces puzzlement to argue that the occurrence versus non-occurrence of an outcome could have a different explanation. STCM has taken an idea that is too often seen as puzzling, and shown that this idea is indispensable.[8]

Other key contributions include a new approach to studying equifinality, i.e., multiple causal paths; distinctive tools for assessing causal interactions; an insistence on the importance of context; a strong commitment to mobilizing case knowledge; and a central emphasis on the interplay of theory and case knowledge that brings together deductive and inductive approaches to gaining new insights.

The field has definitely benefitted from STCM's advocacy of these goals. Indeed, rather than declaring qualitative and quantitative methods to be "two cultures" (Goertz and Mahoney 2012), scholars might instead celebrate the contributions of STCM in advocating these goals for the broader field of methodology—though of course, some quantitative researchers have long promoted many of the same goals.

[5] This recommendation echoes the conclusions of Seawright (2005: 41; 2014); Lucas and Szatrowski (2014); and Collier (2014).

[6] Possibly the single most important innovation in qualitative methods of the past several years is Bennett's (2014) reframing of process tracing, summarized in this symposium, which builds on Humphreys and Jacobs (2013) remarkable new framework for multi-method research.

[7] The term asymmetric causation is also used to characterize a unidirectional causal relation between a given pair of variables. That is not the meaning intended here.

[8] Ironically, the idea of asymmetric causal patterns is more standard in other, very different domains. For example, it is presented in as conventional a source as Fahnestock and Secor's (1982: 132–146; 2nd and 3rd editions 1990 and 2003) textbook for teaching undergraduates good writing skills.

Part 1. Concerns about the Set-Theoretic Method

At the same time, questions have arisen about STCM's tools, and the four contributions to Part 1 of this symposium explore these questions. The present essay raises a number of concerns about these methods, provides an overview of the symposium, and poses questions to suggest future directions for more effectively pursuing these same goals.

The other three authors in Part 1: (a) challenge the idea that set-theory can be justified in part because it reflects the structure of meaning in natural language (Lakoff); (b) argue that an emphasis on constructing well-bounded concepts is a separate matter from embracing the logic and procedures of set theory as a guiding framework for research (Sartori); and (c) raise a number of questions about STCM's approach to causal inference (Tanner).

Set Theory and Natural Language (Lakoff). A recurring theme in discussions of set-theoretic methods is that this approach is compelling in part because it reflects the structure of natural language.[9] The relationship to natural language is addressed here in an interview with the prominent cognitive linguist George Lakoff, whose work is periodically evoked in arguments that justify set-theoretic approaches—including fuzzy-set analysis.[10]

Lakoff dissents from these arguments, drawing on strong evidence that the organization of meaning in natural language is not based on classical categorization, with necessary and sufficient conditions for category membership. While some concepts are well-bounded, prototype theory suggests that a great many are not. Even for those that are well-bounded, prototype theory points to the importance of not reifying these boundaries. This raises serious concerns about the set-theory template.

In addition, Lakoff discusses Zadeh's fuzzy logic, expressing admiration for its application to engineering—yet arguing it is not generally a good match for the structure of meaning in natural language. He also notes the large difference between Zadeh's fuzzy logic and Ragin's procedure for scoring fuzzy-set membership. Lakoff suggests, given the fixed numerical values assigned in Ragin's fuzzy-set scoring, that this analytic procedure should in fact not be considered a fuzzy method.

The Quest for Well-Bounded Concepts (Sartori). Crafting well-bounded concepts has long been a central priority for methodologists and also for applied researchers. Giovanni Sartori is a leading advocate of this practice, and his work is evoked by advocates of set theory and associated systems of logic.[11] For the purposes of social science, Satori insists on classical categorization, based on necessary and sufficient criteria for category membership. However, he rejects the application of set theory as a central technique in qualitative re-

[9] Ragin (2008: especially 38; also 2, 13, 97); Goertz and Mahoney (2012: 11–12, 16–18); Schneider and Wagemann (2012: 7; 2013: 21–22).

[10] See Ragin (2000: 6, 171; 2008: 98); Goertz and Mahoney (2012: 16, 18); Schneider and Wagemann (2013: 21–22).

[11] Ragin (2000: 321, 328; 2008: 98); Goertz and Mahoney (2012: 6, 12, 16, 139, 148).

search. He draws on his long-standing distinction between the unconscious thinker, who fails to reflect on concepts and methods, and the over-conscious thinker, who is counter-productively focused on techniques that may well be more complex than is productive for the task at hand (Sartori 1970: 1033–1040).

Sartori seeks to follow a middle path. For example, in developing his well-known ladder of abstraction, he did make limited use of ideas from logic, and he does think that scholars should be familiar with the tools of logic. However, he rejects the proposal that the set-theoretic approach and associated forms of logic should become a dominant framework, because they may lead the researcher to become bogged down in unproductive techniques.

Tools for Causal Inference (Tanner). Sean Tanner, a scholar of public policy, makes a two-fold argument. Using many examples, he reviews QCA's tools for causal inference and finds them problematic. Tanner also addresses the argument advanced by QCA scholars that their distinctive approach to causal inference is valuable for the study of public policy. Tanner suggests that evaluation research is an area of policy analysis that places especially strong demands on tools of causal inference, and policy evaluation is therefore an appropriate "crucial case" for assessing QCA's value for public policy studies.

Tanner offers a detailed comparison between policy analyses based on QCA, as opposed to studies that follow today's standard norms for policy research. He finds that the conventional studies yield the kind of insights policy analysts urgently need, whereas the QCA analyses offer findings that are too often unhelpful and uninterpretable. The discussion includes such topics as gaining insight into causal interactions; the importance of "net effects thinking," an approach strongly questioned by QCA scholars; the problem that QCA scoring too often yields measurements that are difficult for policy analysts to interpret; the extremely large—and therefore, again, hard to interpret—number of causal paths often yielded by QCA; and the exceedingly small number of cases per causal path, sometimes just one or two.

Throughout, Tanner emphasizes current standards for causal inference that mandate careful choices in making inferences from observational data—and indeed, from all kinds of data. By these standards, he finds QCA to be seriously deficient.

Part 2. Where Do We Go From Here?

Building on these commentaries, the second part of the symposium asks: "Where do we go from here?" Topics include: (a) the challenge of developing simulations that are appropriate for evaluating the stability and validity of findings derived from STCM; (b) alternative procedures for analyzing partial membership in categories; (c) contrasting approaches to the study of interactions among different combinations of explanatory factors; and (d) a new approach to process tracing that moves beyond earlier criteria of necessity and sufficiency for evaluating causal inference.

Evaluating Simulations (Krogslund and Michel). The use of computer simulations to evaluate alternative methods is now standard across the social sciences. Correspondingly, the most important area of assessment and innovation in discussions of STCM currently involves simulations that evaluate the stability and validity of findings. This topic merits close attention here.

STCM employs a complex set of algorithms,[12] and a growing number of studies have raised concerns that these algorithms are highly sensitive to small changes in measurement decisions and to shifts in the parameters that must be set for causal inference.[13] A recurring finding has been a tendency to generate false positives.

From within the STCM tradition, Schneider and Wagemann's overview of simulations and robustness is more encouraging, but they conclude with great caution: "QCA is not vastly inferior to other comparative methods in the social sciences" (2012: 294). This is faint praise, given the numerous, sharp critiques of the stability of findings based on conventional quantitative analysis—obviously a key method of comparison.

Considerable attention is being devoted to the crucial question of how to design tests that reproduce the algorithms employed in these methods, and Thiem (2013), for example, has made a key contribution. Readers should be alerted to forthcoming debates that are unfortunately not yet available (including to this author) at the time of this newsletter's publication.[14]

The jury is definitely still out in terms of assessing specific simulation tests—at the same time that the cumulative evidence raises very strong concerns about STCM's tools. These concerns are reinforced by the fact that STCM has not incorporated into its analytic procedures a recognition of the recent transformation of thinking about causal inference, as practiced in all social science methods, which is discussed—as noted above—by Tanner and by Collier (2014).

[12] Algorithms are understood as systematized procedures for making calculations, often implemented with computer software. QCA's ensemble of algorithms includes, for example, procedures that address contradictions, logical remainders, minimization, sufficiency scores, minimum frequency, consistency, coverage, and the probabilistic criteria for causal inference.

[13] Hug 2013; Seawright 2013, 2014; Kurtz 2013; Krogslund, Choi, and Poertner 2013; Krogslund and Michel 2014; Lucas and Szatrowski 2014.

[14] Some time ago, Ragin and Rihoux (2004: 22–24) argued that an earlier evaluation based on hypothetical data was not suitable to QCA. Readers should watch for important, forthcoming exchanges. Thus, an important debate has been generated by Lucas and Szatrowski's (2014) simulations. Ragin (2014) has written a commentary on their article, and they in turn have responded to his commentary. Given the norms of the journal *Sociological Methodology* where this exchange will appear, these comments are not available to other contributors to that symposium until the time of their actual publication. In addition, Thiem's (2014) commentary on Hug (2013) is scheduled to be included in the next issue of the present newsletter, accompanied by a response from Hug. When these publications become available, they should and will play an important role in these debates.

Krogslund and Michel's contribution to this symposium seeks to advance this important search for appropriate simulations by evaluating results from a "drop-one" sensitivity test of QCA. Their initial finding from this test—intriguingly—inverts Arend Lijphart's (1971: 686) traditional formulation of the "many variables, small-N problem" in comparative research. Thus, Krogslund and Michel's findings suggest QCA might, ironically, have a "fewer variables, larger-N" problem. That is to say, findings appear more unstable, to the degree that the analyst focuses on more cases and a smaller number of explanatory variables. This counter-intuitive result leads them to scrutinize the properties of this test, as well as its appropriateness to the analytic procedures of QCA—with a particular emphasis on how logical remainders (i.e., empty rows in the truth table) are treated.

Simulation tests will be crucial in the ongoing evaluation of QCA, and Krogslund and Michel illustrate the kind of painstaking, fine-grained analysis needed to adequately assess the appropriateness of specific tests.

Measuring Partial Membership in Categories (Elkins). Fuzzy-set analysis was introduced as a tool for measuring partial membership in categories, and Zachary Elkins' contribution evaluates alternative approaches to the study of partial membership. He raises some of the same concerns advanced in this symposium by Lakoff (2014) and Tanner (2014) about the ambiguities of scoring fuzzy-sets. On the basis of these concerns, Elkins advocates attention to three issues: the conceptual structure of the categories, homogeneity within categories, and degree of membership.

Elkins focuses on a substantive example that is highly salient to the present discussion: the comparative study of constitutions. The categories of presidentialism, semi-presidentialism, and parliamentarism are widely recognized and—presumably—extremely well defined, yielding an excellent opportunity to explore ideas about full membership and partial membership. Elkins applies three scaling techniques to his data—MIMIC Modeling (Multiple-Indicators Multiple Causes), similarity-based measures and latent-class analysis—using them to assess the structure and heterogeneity of the categories and degrees of membership. He suggests that these methods have many advantages over fuzzy-set scoring and they yield important new insights into the complexities of these extremely well-known categories of legislative-executive relations.

Studying Interactions (Braumoeller). A major concern of social science methodology is with how causal patterns differ when distinct combinations of causes interact. This topic is addressed under various rubrics, including the study of interactions and of contextual effects. Concern with causal patterns such as these is a hallmark of the qualitative tradition. Bear Braumoeller's contribution systematically compares four approaches to the study of interactions: QCA's focus on causal combinations, interaction terms in regression analysis, stochastic frontier modeling, and Boolean logit.

These methods differ in their approach to measurement, the role of coefficients, the treatment of thresholds, conceptualization of the interactions themselves, and data requirements. Clearly, none of these models is the "correct" one, and Braumoeller's contribution is an excellent point of departure for understanding the opportunities, limitations, and trade-offs that arise in studying interactions. It establishes a broad agenda for future work on this crucial topic.

Process Tracing: Beyond the Criteria of Necessity and Sufficiency (Bennett). Process tracing is a fundamental tool relevant to all forms of research, both qualitative and quantitative, and Andrew Bennett's contribution is a major stride forward.[15] A central goal of ongoing research on process tracing has been to provide criteria for evaluating the four process-tracing tests proposed some time ago by Van Evera (1997: 31-32): straw-in-the-wind, hoop, smoking-gun, and doubly-decisive tests.

One approach to mapping the relationship among these tests, adopted by Bennett (2010) and Collier (2011), had been to differentiate according to whether each test is necessary and/or sufficient for affirming a given causal inference. Goertz and Mahoney (2013: 279) have noted this approach as an example of applying set theory and related forms of set logic to qualitative methods.

However, Bennett has now moved the process-tracing literature well beyond this approach, based on his application of Bayesian analysis. In Bennett's new formulation, the overall goal is the same: evaluating the power of process-tracing evidence for testing a given hypothesis. The Bayesian approach systematizes insights into this probative power based on three criteria: (1) whether positive evidence is found, (2) the researcher's prior confidence that the hypothesis is correct, and (3) the likelihood ratio, i.e. the odds of finding positive evidence if the explanation is correct versus those of finding the same evidence even if the explanation is false.

In this framework, the four traditional tests can be situated within a spectrum of possibilities. This spectrum is established based on the degree to which (a) finding or not finding the evidence has (b) a modest or strong effect on (c) the posterior assessment of whether the hypothesis is supported.

Bennett argues that the ideas of necessary and sufficient are superseded because they are too categorical, and he advocates this more flexible approach. In a sense, the differentiation of four process-tracing tests has also been superseded, and these tests can now most usefully be seen simply as useful benchmarks within this continuous spectrum of alternative inferences.

With the Bayesian method, a key question is whether researchers should "fill in the numbers," using the Bayesian algorithms to actually make calculations; or instead employ the method as a useful heuristic for reasoning about process tracing. This second version would correspond to McKeown's (2004: 158–162) idea of "folk Bayesianism." Bennett is open to both alternatives—and in either case, a central point is that the categorical framing of necessity and sufficiency as a basis for evaluating process-tracing tests is replaced by a framing based on the idea of continuous gradations.

It should be added that scholars identified with STCM

[15] More broadly, see the major new book on process tracing by Bennett and Checkel (2014).

have also recently introduced new perspectives on process tracing, and this method promises to be a key area of future innovation.[16]

Some Further, Exploratory Questions

The essays in this symposium call for a fundamental rethinking of the tools employed by the set-theoretic comparative method. This section poses several exploratory questions—some of which correspond to innovations currently being developed by STCM scholars—that may be useful points of reference in this rethinking. The first three questions focus on measurement, the others causal inference.

Is fuzzy-set scoring really fuzzy? Lakoff observes that given the fixed values assigned in fuzzy-set scoring, the method is in fact not fuzzy. He notes the contrast with Zadeh's scoring procedure, in which the assigned values are themselves fuzzy, rather than fixed.

Is fuzzy-set scoring viable? Lakoff, Tanner, and Elkins are all concerned about the lack of clear standards for assigning scores, and hence for problems with the interpretability of scores. For example, if one finds the initial designation of full membership in the set to be ambiguous, then the rest of the scale becomes ambiguous. Tanner is also concerned that even with the fuzzy-set version, the method yields aggregated causal paths that lose a great deal of information.

These concerns point to a further question: do the standard indicators on which fuzzy-set scores are often based convey more useful information than the fuzzy-set scores themselves? These indicators have the advantage of being formulated in terms of familiar and more readily interpretable units of measurement.

Is fuzzy-set scoring necessary? Correspondingly, might it be preferable if STCM scholars made greater use of standard indicators, rather than fuzzy-set scoring? If one accepts the possibility that fuzzy sets are in fact not fuzzy, then perhaps not a great deal would be lost. Analysts would retain the more readily intelligible units of measurement. At the final step in causal inference, when the fuzzy-set findings are dichotomized for entry into the truth table, STCM scholars could draw on the insights gained in the course of the analysis to make what might be better-informed judgments about establishing the cut-points for dichotomization.

This approach embraces Ragin's (2000: 171) recommendation that recoding fuzzy membership scores in the course of the analysis should be considered standard practice. In addition, the important goal of eliminating "irrelevant variation" (Ragin 2000: 161–63) would be achieved through the dichotomization introduced at this final step in the process—again, hopefully guided by these better-informed judgments that emerge in the course of the study.

A final point: In addition to encouraging greater use of standard indicators, this framework could incorporate a broader set of tools. For example, Elkins' methods for analyzing partial membership in categories might be considered just as appropriate as fuzzy-set scoring for these STCM applications.

What are the next steps in analyzing interactions? Tanner expresses the concern that QCA's approach to studying interactions routinely does not yield productive insights, and Braumoeller takes a large stride toward extending the discussion by noting major tradeoffs among four alternative models of interactions, including QCA.

Braumoeller's line of analysis must be developed much further—for example, analysts might wish to add a fifth approach. Tanner's examples include the compelling analysis of an interaction that relies on a simple two-by-two table, involving welfare interventions for teen-aged mothers. Tools for analyzing standard cross-tabulations are much neglected in today's social science, and obviously there is a specific "algorithm" that one can follow for such analysis. This algorithm might be added as a fifth model for studying interactions.

Have the algorithms and software taken over, and has case knowledge been eclipsed? This symposium has suggested the answer is yes, and this question merits continuing attention. Consider the many empirical studies using QCA that are analyzed by Tanner and by Krogslund and Michel—along with dozens of additional empirical articles using the method. The concern does indeed arise that the method is in effect reduced to the algorithms—which are all too readily applied using QCA software. The intensive use of case knowledge is often not in evidence.

If we look at the trajectory of other innovative methods, we see that the widespread availability of software can be both a blessing and a curse. In the case of structural equation modeling, it opened a Pandora's Box of bad applications (Steiger 2001). One worries that the same distortion of the method has occurred with QCA.

How should one think about problems of stability and validity of findings that emerge in simulations? Error and the "DGP." As scholars conduct more simulation tests, it is essential to ask why these tests often reveal problems of stability and validity. One possibility is that STCM lacks adequate tools for dealing with a number of issues, including measurement error, problems of model specification, and potentially a random element in the data.

In evaluating these issues of the stability and validity of findings, STCM scholars should consider a concept that is crucial to methodological discussions today: the underlying "data-generating process" (DGP), which focuses attention on what is now the standard insight that causal assessment involves reaching conclusions about the DGP on the basis of the particular set of observed values. This is a useful way of bringing into sharp focus the daunting challenges of causal inference. Some simulation tests specifically evaluate STCM within this DGP framework (Krogslund and Michel 2014), and these tests reinforce concerns about the stability and validity of findings. Additional work along these lines will make a large contribution to evaluating STCM's analytic tools.

Two more questions about the stability and validity of

[16] See Mahoney (2012); Schneider and Rohlfing (2013); Rohlfing and Schneider (2013); Beach and Pedersen (2013); Rohlfing (2014).

findings also merit attention.

Is there an asymmetry in the treatment of false negatives and false positives? STCM has procedures such as its probabilistic criteria to help guard against false negatives. Is it possible that parallel procedures to guard against false positives are lacking? Regarding false negatives: in the real world, given problems that include measurement error, a true causal relationship of necessity and/or sufficiency may be present, yet it might be imperfectly reflected in the data. For example, if the analysis employs a two-by-two table, some cases might be located in the "wrong" cells, vis-à-vis the true causal pattern. It appears that STCM has procedures for addressing this problem, thereby guarding against a potential false negative.

However, it is not clear that there is a procedure for addressing the opposite problem, which could yield a false positive. Thus, cases might be located in the "right" cells for inferring necessity and/or sufficiency, yet in an analysis free of error they might in fact be located in the "wrong" cells. The resulting inference runs the risk of being a false positive.

The question of whether STCM has appropriate tools for addressing both false negatives and false positives requires much further discussion.

How should the empty rows in the truth table—the logical remainders—be addressed? For scholars who are not part of the STCM tradition, the treatment of empty rows—i.e. combinations of conditions not found in the empirical cases—is puzzling. The problem of "limited diversity" addressed in STCM is definitely important, and the counterfactual reasoning that underlies causal inference routinely involves empty rows. Yet STCM appears, overall, problematic in meeting the challenge posed by standard norms of causal inference; and employing an extremely complicated analytic procedure to address what may well not be the most compelling aspect of this deficit seems questionable.

The truth table, a foundation of QCA, is a logical construct, and as a logical construct it encompasses all possible combinations of explanatory conditions that could be matched with both the occurrence and non-occurrence of the outcome. As is widely noted, with additional explanatory conditions this leads to an exponential increase in the overall number of rows, and also to a dramatic increase in empty rows. This yields a cascade of complications for the method, along with the need for complex algorithms to address these complications.

In the actual practice of STCM, the truth table might more usefully be treated not as a logical construct, but as a valuable form of data display. Correspondingly, the focus on the empty paths could be dropped from the method.

In the field of comparative-historical analysis—for which STCM is intended to have great value—a focus on empty paths would be non-standard. For example, in *Shaping the Political Arena* (R. Collier and D. Collier 1991), we could not possibly have been able to—or wanted to—address a large number of empty rows. Many books have a concluding chapter that somewhat speculatively places the cases analyzed in a wider comparative perspective—including potentially some comments on empty rows. This is valuable, and it strengthens causal inference. But elaborate attention to empty rows is emphatically not a cornerstone of the comparative-historical method.

If the truth table were simply treated as a valuable data display, perhaps the concern with empty rows could simply be dropped, and attention might usefully focus on other limitations of STCM's procedures for causal inference.

To reiterate, these are exploratory questions that seek to advance the discussion. Some questions correspond to innovations currently being developed by STCM scholars—although the findings of the present essay suggest that the overall goals of the method may be more effectively served by turning to different tools.

Conclusion: Restoring Ragin's Dialogue between Ideas and Evidence

Taken together, these questions point to the need for a fairly drastic reevaluation of the tools employed in the set-theoretic method. To place this reevaluation in perspective, a concluding point should be made about a central foundation of the method, which grows out of the work of Charles Ragin.

One of Ragin's fundamental scholarly contributions is his conception of social research as a dialogue between ideas and evidence. This conception is important for QCA, but very crucially also in his many non-QCA books and articles on methods (Ragin and Zaret 1983; Ragin and Becker 1989, 1992; Ragin 1994, 2004; Ragin and Amoroso 2010). This trajectory in his work can readily be seen as a creative extension of the long sociological tradition that includes, for example, the Lazarsfeld elaboration model and the constant comparative method of grounded theory. This tradition remains a cornerstone of good research.

As noted, in much of the applied work using QCA, this component may well have disappeared. It seems that the role of case knowledge—and the dialogue between ideas and evidence—is being eclipsed by the algorithms and the computer software. Setting aside the algorithms—combined with concentrating on case studies and process tracing, as is advocated here—would bring the focus back to Ragin's larger contribution.

Next Steps: Following through on this Symposium

The *Qualitative and Multi-Method Research* newsletter is committed to publishing in its next issue a comment on this symposium from the perspective of STCM, a response from the standpoint of the symposium, a comment on the recent Hug (2013) article, and a response from Hug. This should be a valuable exchange, and we strongly welcome it.

Regarding this evolving discussion, an observation should be made about a direction the debate hopefully will *not* take. In earlier exchanges, the skeptical evaluation of QCA by two commentators was challenged as reflecting the limited perspective and "defensive reactions" of quantitative researchers.[17] In another exchange, a skeptical evaluation by a third

[17] See Ragin and Rihoux's (2004: 22) comments on the evaluations by Lieberson and Seawright.

commentator was dismissed as "what one would expect a quantitative researcher to believe."[18]

These responses by STCM scholars took the discussion down the wrong path, particularly because all three of these commentators are specifically well-known as strong critics of conventional quantitative methods.[19] In the spirit of this newsletter's commitment to multi-method research, we have sought to meet a key standard: to the extent possible and feasible, contributors should have a broad view of methodology that transcends limitations such as these.

In that framework, we greatly look forward to the continuing discussion.

References

Beach, Derek, and Rasmus Brun Pedersen. 2013. *Process-Tracing Methods: Foundations and Guidelines*. Ann Arbor: University of Michigan Press, 2013.

Bennett, Andrew. 2010. "Process Tracing and Causal Inference." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, eds. Henry E. Brady and David Collier. Lanham, MD: Rowman and Littlefield, 207–219.

Bennett, Andrew. 2014. "Process Tracing with Bayes: Moving beyond the Criteria of Necessity and Sufficiency." *Qualitative & Multi-Method Research* 12 (1): 46–51.

Bennett, Andrew and Jeffrey Checkel. 2014. *Process Tracing in the Social Sciences: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press.

Brady, Henry E. 1995. "Doing Good and Doing Better." *The Political Methodologist* 6 (2): 11–19.

Brady, Henry E. and David Collier, eds. 2004, 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman and Littlefield.

Braumoeller, Bear. 2014. "Analyzing Interactions: Four Alternative Models." *Qualitative & Multi-Method Research* 12 (1): 41–46.

Collier, David. 2011. "Understanding Process Tracing." *PS: Political Science and Politics* 44 (4): 823–830.

Collier, David. 2014. "QCA Should Consider Abandoning the Algorithms." *Sociological Methodology* 44: Forthcoming.

Collier, Ruth Berins and David Collier. 1991. *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. Princeton: Princeton University Press.

Elkins, Zachary. 2014. "Measuring Partial Membership in Categories: Alternative Tools." *Qualitative & Multi-Method Research* 12 (1): 33–41.

Fahnestock, Jeanne, and Marie Secor. 1982. *A Rhetoric of Argument*. 2nd and 3rd eds, 1990 and 2003. New York: McGraw-Hill.

Goertz, Gary, and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, New Jersey: Princeton University Press.

Goertz, Gary and James Mahoney. 2013. "For Methodological Pluralism: A Reply to Brady and Elman." *Comparative Political Studies* 46 (2): 278–285.

Hug, Simon. 2013. "Qualitative Comparative Analysis: How Inductive Use and Measurement Error Lead to Problematic Inference." *Political Analysis* 21 (2): 252–265.

Humphreys, Macartan and Alan Jacobs. 2013. "Mixing Methods: A Bayesian Unification of Qualitative and Quantitative Approaches." Paper presented at the Annual Meeting of the American Political Science Association, Chicago, IL.

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

Krogslund, Chris and Katherine Michel. 2014a. "Evaluating Set-Theoretic Methods: Finding Appropriate Simulations." *Qualitative & Multi-Method Research* 12 (1): 25–33.

Krogslund, Chris and Katherine Michel. 2014b. Recovering the Data-Generating Process: How Well Do Set-Theoretic Methods Perform? Revised version of a Paper Presented at the 2014 annual meeting of the Midwest Political Science Association, April 3-6, Chicago IL.

Krogslund, Chris, Donghyun Danny Choi, and Matthias Poertner. 2013. "Fuzzy Sets on Shaky Ground: Parameter Sensitivity and Confirmation Bias in fsQCA." Revised version of a Paper Presented at the Annual Meeting of the American Political Science Association, Chicago.

Kurtz, Marcus. 2013. "The Promise and Perils of Fuzzy-set/Qualitative Comparative Analysis: Measurement Error and the Limits of Inference." Presented at the Annual Meeting of the American Political Science Association, Chicago.

Laitin, David D. 1995. "Disciplining Political Science." *American Political Science Review* 89 (2): 454–456.

Lakoff, George. 2014. "Set Theory and Fuzzy Sets: Their Relationship to Natural Language. Interview with George Lakoff, Conducted by Roxanna Ramzipoor." *Qualitative & Multi-Method Research* 12 (1): 9–14.

Lieberson, Stanley. 1985. *Making It Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.

Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65 (3): 682–693.

Lucas, Samuel R. and Alisa Szatrowski. 2014. "Qualitative Comparative Analysis in Critical Perspective." *Sociological Methodology* 44: Forthcoming.

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

Mahoney, James. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods & Research* 41 (4): 570–597.

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

Ragin, Charles C. 1994. *Constructing Social Research*. Thousand Oaks, CA: Pine Forge Press.

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

Ragin, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.

Ragin, Charles C. 2014. "Lucas and Szatrowski in Critical Perspective." *Sociological Methodology* 44: Forthcoming.

Ragin, Charles C. and Lisa M. Amoroso. 2010. *Constructing Social Research: The Unity and Diversity of Method*. 2nd ed. Los Angeles: Sage Publications.

Ragin, Charles C. and Howard S. Becker. 1989. "How the Microcom-

[18] See Goertz and Mahoney's (2013: 279–80) response to the evaluation by Brady.

[19] Regarding the critiques of conventional quantitative methods by these three authors, see Lieberson (1985); Brady (1995), Brady and Collier (2004); and for Seawright, Chapters 1, 2, 9, and 10 in Brady and Collier (2004), and subsequently Seawright (2007). It also merits note that to the extent that scholars might be classified as "quantitative" researchers, they routinely also have a strong background in mathematics and logic—tools essential to these discussions.

puter Is Changing Our Analytic Habits." In *New Technology in Society: Practical Applications in Research and Work*. (New Brunswick, NJ: Transaction Publishers), 47–55.

Ragin, Charles C. and Howard S. Becker. 1992. *What Is a Case? Exploring the Foundations of Social Inquiry*. Cambridge: Cambridge University Press.

Ragin, Charles C. and Benoît Rihoux. 2004. "Qualitative Comparative Analysis (QCA): State of the Art and Prospects." *Qualitative Methods* [subsequently renamed *Qualitative and Multi-Method Re search*] 2 (2): 3–13, 22–24.

Ragin, Charles and David Zaret. 1983. "Theory and Method in Comparative Research: Two Strategies." *Social Forces* 61 (3): 731–754.

Rohlfing, Ingo. 2013. "Comparative Hypothesis Testing Via Process Tracing." *Sociological Methods & Research*. Online First DOI: 10.1177/0049124113503142.

Rohlfing, Ingo and Carsten Q. Schneider. 2013. "Improving Research on Necessary Conditions: Formalized Case Selection for Process Tracing after QCA." *Political Research Quarterly* 66 (1): 220–230.

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

Sartori, Giovanni. 2014. "Logic and Set Theory: A Note of Dissent." *Qualitative & Multi-Method Research* 12 (1): 14–15.

Schneider, Carsten Q. and Ingo Rohlfing. 2013. "Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research." *Sociological Methods & Research* 42 (4): 559–597.

Schneider, Carsten Q. and Claudius Wagemann. 2013. "Fuzzy Sets Are Sets—A Reply to Goertz and Mahoney." *Qualitative & Multi-Method Research* 11 (1): 20–23.

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. 2005. "Assumptions, Causal Inference, and the Goals of QCA." *Studies in Comparative International Development* 40 (1): 39–42.

Seawright, Jason. 2007. "Democracy and Growth: A Case Study in Failed Causal Inference." In *Regimes and Democracy in Latin America: Theories and Methods*, ed. Gerardo L. Munck. Oxford: Oxford University Press, 179–198.

Seawright, Jason. 2013. "Warrantable and Unwarranted Methods: The Case of QCA." Presented at the Annual Meeting of the American Political Science Association, Chicago.

Seawright, Jason. 2014. "Limited Diversity and the Unreliability of QCA." *Sociological Methodology* 44: Forthcoming.

Steiger, James H. 2001. "Driving Fast in Reverse: The Relationship between Software Development, Theory, and Education in Structural Equation Modeling." *Journal of the American Statistical Association* 96 (453): 331–338.

Tanner, Sean. 2014. "Evaluating QCA: A Poor Match for Public Policy Research." *Qualitative & Multi-Method Research* 12 (1): 15–25.

Thiem, Alrik. 2013. "Membership Function Sensitivity of Descriptive Statistics in Fuzzy-Set Relations." *International Journal of Social Research Methodology*. Online First DOI: 10.1080/13645579.2013.806118.

Thiem, Alrik. 2014. "Data and Measurement Error in Qualitative Comparative Analysis: An Extended Comment on Hug (2013)." *Qualitative and Multi-Method Research* 12 (2): Forthcoming.

Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.

Wedeen, Lisa. 2014. "Letter from the Section President." *Qualitative and Multi-Method Research* 12 (1): 1–2.