

This paper is a work in progress. Comments are eagerly sought, as is tolerance for mistakes, stylistic infelicities, and other (hopefully transitory) deficiencies.

# **Questioning Set-Theoretic Comparative Methods**

## **Admirable Goals, Problematic Tools, Better Options**

**David Collier**

**Thad Dunning**

Department of Political Science  
University of California, Berkeley  
November 13, 2014

<b>1. Introduction</b> .....	<b>2</b>
<b>2. Admirable Goals</b> .....	<b>4</b>
<b>3. Concepts and Measurement</b> .....	<b>6</b>
<b>3a. Summary of Basic Framework</b> .....	<b>6</b>
<b>3b. Concerns about Concepts and the Set-Theoretic Framing</b> .....	<b>8</b>
<b>3c. Concerns about Measurement</b> .....	<b>11</b>
<b>4. Causal Inference</b> .....	<b>16</b>
<b>4a. Summary of Basic Framework</b> .....	<b>16</b>
<b>4b. Concerns about Causal Inference</b> .....	<b>23</b>
<b>5. Methodological Path Dependence?</b> .....	<b>36</b>
<b>6. Better Options</b> .....	<b>39</b>
<b>6a. Innovation in Traditional Tools</b> .....	<b>40</b>
<b>6b. Algorithmic-Based Tools</b> .....	<b>44</b>
<b>6c. Assessing Mechanisms</b> .....	<b>47</b>
<b>7. Toward a Conclusion</b> .....	<b>49</b>
<b>Bibliography</b> .....	<b>51</b>

## 1. Introduction

The long-standing challenge of systematizing qualitative research has recently been addressed by innovative work on the set-theoretic comparative method (STCM).<sup>1</sup> This approach is strongly identified with Charles Ragin's "Qualitative Comparative Analysis" (QCA),<sup>2</sup> and that abbreviation is used occasionally below when the discussion focuses specifically on Ragin's contributions. STCM maintains that set theory successfully organizes the tasks of conceptualization, measurement, and causal assessment—thereby offering a new framework for designing research. This framework has received wide attention and merits careful and respectful attention

The present article endorses STCM's worthy *goals*. This method was built on a well-focused critique of conventional quantitative research, and it addresses issues too often neglected in the quantitative tradition: concepts, case knowledge, context, and causal complexity. STCM has been an important force in advocating analytic breadth, especially in the field of comparative and international studies. It has definitely not been the only line of research that pursued these priorities, yet STCM was and still is noteworthy for offering an integrated approach to addressing them.

However, the discussion below questions the STCM's tools for pursuing these important goals. (1) A central justification for this method is the claim that set theory reflects the structure of meaning in natural language and qualitative analysis appears questionable. (2) STCM too emphatically rejects other approaches to measurement and

---

<sup>1</sup> Ragin (1987, 2000, 2008); Rihoux and Ragin (2009). Goertz and Mahoney (2012) and Schneider and Wagemann (2012) are also major statements in this set-theoretic tradition.

<sup>2</sup> QCA is understood to encompass the fuzzy-set version—fsQCA—as well as csQCA and mvQCA.

overstates the novelty of its own procedures. (3) By the norms of the Zadeh tradition of fuzzy logic, fuzzy sets in STCM are not fuzzy measurement, but rather are equal-interval, linear scales.

With regard to causal inference, (4) STCM's basic tests often fail to yield the kind of insights routinely sought by many scholars—for example, findings based on real-world units of measurement, forms of interactions other than those that are the focus of STCM, and tests of the relative importance of different predictor variables. (5) The argument that the causal inferences derived from process tracing and case-based analysis inherently involve necessary and/or sufficient conditions is questionable. (6) STCM presents intriguing innovations in the analysis of necessity and sufficiency—for example, the ideas of “usually” and “almost always,” as well as “probabilistically” necessary/sufficient. Yet scholars outside this tradition may find that these innovations dilute the deterministic framework that made this approach attractive in the first place. (7) In simulation tests, findings often prove to be unstable and/or invalid; and given STCM's analytic procedures, these simulation results are not surprising. (8) The recent rethinking of causal inference in the social sciences has not been adequately engaged STCM. This method needs to recognize the major obstacles to achieving its ambitious inferential goals.

By thus embracing STCM's goals, but questioning the tools, this evaluation follows a middle path—in contrast to the polarized alternatives posited, for example, in Elman's review of Goertz and Mahoney (2012). Elman states that this book will be

a major force in helping methodologists to clarify and strengthen their own positions, whether in opposition to or filling the gaps left by Goertz and Mahoney. (Elman 2013, 275)

The present article situates itself between Elman's stark alternatives of opposition and filling gaps.

Following this introduction, Part 2 reviews the goals advocated by STCM. Parts 3a, 3b, and 3c introduce the basic STCM framework for concepts and measurement, raise questions about the set-theoretic framing of concepts, and explore concerns about measurement.

STCM's approach to causal inference is summarized in Part 4a, and Part 4b poses questions about this framework. Part 5 introduces an historical perspective by noting the longer trajectory of critiques concerned with the narrowness of conventional quantitative methods. Part 6 then considers better tools, some of which are traditional methods that have recently been strengthened by new innovations. Based on these tools, the overall goals of STCM regarding context, concepts, and complexity can be pursued more effectively, but preferably not in a set-theoretic framework.

## **2. Admirable Goals**

The goals of STCM, as noted above, center on concepts, case knowledge, context, and causal complexity. As noted, STCM's goals were formulated to address the analytic narrowness of conventional quantitative methods. This narrowness has been criticized from diverse perspectives—for example, in the work of Christopher Achen discussed below. But STCM has unquestionably made an important contribution to sustaining this critique.

Conventional quantitative methods are seen by STCM proponents as: (1) naïvely variable-oriented and insufficiently case-oriented; (2) based on an inadequate understanding of “variance” and a failure to distinguish between relevant and irrelevant variation; (3) relying on a linear-additive model that does not incorporate contextual effects, interactions, and other interdependencies among explanatory factors—thus

neglecting causal complexity; (4) and failing to gain leverage in addressing these limitations by building on an iterated examination of theory and case knowledge.<sup>3</sup> This calls for a “dialogue between ideas and evidence” advocated by Ragin in his QCA (and non-QCA) methodological writing.<sup>4</sup>

Against this backdrop, STCM proposes the following, alternative agenda. (1) A focus on *concepts*, which are one side of this dialogue between ideas and evidence. (2) Attention to *case knowledge*, as a foundation of good research and as the other side of the dialogue between ideas and evidence. (3) Concern with the *context* of an event or action, as essential to studying it adequately. The standard term “contextual effects” is rarely employed in this literature,<sup>5</sup> but the underlying idea is central to the method. (4) Recognition of *causal complexity*, particularly *equifinality*, *interactions*, and *asymmetric causation*. Equifinality—the idea of multiple paths to a given outcome—is familiar in various research traditions, but is well worth emphasizing. Interactions have likewise received wide attention, and STCM offers approach focused on combinations of conditions.

The idea of asymmetric causation<sup>6</sup>—to reiterate, that the occurrence versus non-occurrence of an outcome may have different explanations—merits special comment here. Thus, a blocking cause ensures the non-occurrence of an outcome; by contrast, a triggering cause ensures the outcome. A standard example is the long trajectory of writing on prerequisites, for example, of democracy. The absence of a prerequisite is a

---

<sup>3</sup> See Ragin 1987: xii, 10, 11, 25, 27; Ragin 2000: 4, 6, 35, 57, 87, 313; Ragin 2008: 7, 33, 74, 77, 83; Rihoux and Ragin 2009: 9, 14, 25, 160.

<sup>4</sup> On the dialogue between ideas and evidence, see Ragin and Zaret 1983; Ragin and Becker 1989, 1992; Ragin 1994, 2004; Ragin and Amoroso 2010.

<sup>5</sup> But see Ragin 2000: 52.

<sup>6</sup> The term asymmetric causation has a second meaning—i.e., a unidirectional causal relation between a given pair of variables (Lieberson 1987: chap. 4). This is not the meaning intended here.

*blocking* cause. The presence of the prerequisite does not by itself produce the outcome, and other factors are relevant for causing it to occur.

For many political scientists, it may initially produce skepticism to argue that the occurrence versus non-occurrence of an outcome could have a different explanation. To be sure, the novelty of this idea can be oversold. From a different perspective, the presence of a triggering cause merely *increases* (the prevalence of) an outcome, while the presence of a blocking cause *decreases* it, both relative to the absence of the cause. Yet STCM has indeed pointed to the value of an idea seen by some scholars as puzzling, and STCM has made a contribution by calling attention to it.

### **3. Concepts and Measurement**

#### **3a. Summary of Basic Framework**

A key focus of STCM is on concepts. This approach argues that set theory reflects the conceptual structure of natural language and of qualitative research. The set-theoretic approach is presented not simply as a methodological recommendation, but is justified because it is seen as a valuable description of ordinary usage and verbal theory.<sup>7</sup>

Thus, Ragin suggests that “almost all social science theory is verbal and, as such, is formulated in terms of sets and set relations” (2008: 13; also 97). He notes that, “unfortunately, social scientists have been slow to recognize this fact” (2008: 97). Goertz and Mahoney make basically the same argument.<sup>8</sup> Given that set theory is a

---

<sup>7</sup> Ragin 2008: 2 and passim; Goertz and Mahoney 2012: 3, 18.

<sup>8</sup> Goertz and Mahoney equate logic and set theory (2012: 12; 16, n. 1), arguing that:

When qualitative scholars formulate their theories verbally, they quite naturally use the language of logic. We refer to this as the “Monsieur Jourdain” nature of the relationship between qualitative scholarship and logic. Qualitative researchers speak the language of logic, but often are

framework for analyzing clearly demarcated clusters of objects or elements, the structure of meaning in natural language and qualitative research is thus seen as inherently well-bounded.

Another central focus is on the treatment of cases and comparisons. STCM calls for attention to “kinds of cases,” understood in set-theoretic terms, whereas quantitative research is seen as concerned with “relationships between variables.” Correspondingly, “case-oriented methods are viewed as holistic—they treat cases as whole entities and not as collections of parts (or as collections of scores on variables)” (Ragin 2008: 101). These contrasts are thoroughly explored in Ragin’s discussions of case-oriented versus variable-oriented research, and when the distinction was first introduced by Ragin (1987), it was a valuable wake-up call for a number of scholars.

The initial version of STCM scored cases in terms of dichotomous conditions, with members of a set scored as 1, and non-members as 0. The fuzzy-set version adds the idea of gradations, but is still anchored in a conception of well-defined set membership. Fuzzy-set scoring retains the overall scores of 1 and 0; identifies a crucial cross-over point of maximum ambiguity between membership and non-membership (0.5); and incorporates further gradations between 0 and 0.5, as well as between 0.5 and 1. Notwithstanding the gradations in fuzzy sets, the binary framing is fundamental. In the final step of causal inference, findings are reported in terms of causal paths or “causal recipes” which are defined dichotomously—though the idea of gradations is retained, in that cases may have partial membership in specific paths.

---

not completely aware of that fact. To systematically describe qualitative research practices, however, it is necessary to make explicit and formalize this implicit use of logic. (Goertz and Mahoney 2012: 11)

These authors explicitly make this same argument not only for qualitative research, but also for natural language (2012: 12, 17–18, and passim).

The calibration—i.e., scoring—of the fuzzy sets is presented as strongly anchored in the theoretical ideas entailed in sets (Ragin 2000: 65; 2008: 33). Ragin posits a sharp contrast with other measurement traditions, arguing that “measurement, as practiced in the social sciences today, remains relatively haphazard and unsystematic” (2008: 74). An example of theoretical anchoring with fuzzy sets is the elimination of “irrelevant variation,” which, it is argued, is too often included in standard measurement practices (Ragin 2000: 6; 2008: 33). For example, once a country is shown to have membership in the set of wealthy countries, variations within that category are seen as irrelevant. On the fuzzy-set scale, wealthy countries therefore all receive a membership score of 1.

Assignment of scores is carried out in two ways. With the *direct method*, standard quantitative measures of relevant concepts are transformed and mapped onto fuzzy-set scores. With the indirect method, the researcher relies on knowledge of cases and context to assign scores to each case (2008: 85).

### **3b. Concerns about Concepts and the Set-Theoretic Framing**

*Natural Language.* The claim that the structure of meaning in natural language is set-theoretic—which to reiterate is used as a justification for the adoption of set theory as an overall method—has long been challenged. Lakoff and Johnson (1980: 71) argued some time ago, based on well-established experimental evidence, that “people characterize objects not in set-theoretical terms, but in terms of prototypes and family resemblances.” Lakoff’s (1987) subsequent analysis of prototype theory repeatedly underscores the contrast with classical categorization, which is central to set theory.

According to the theory of radial categories (Lakoff 1986: chap. 6), an elaboration of prototype theory, extensions of categories routinely do not branch out from the prototype in a linear pattern. Instead, they extend in multiple ways and directions. The



idea of membership and non-membership in sets posits a linear pattern, and is therefore not helpful for reasoning about this conceptual structure. Lakoff (2014: 13) argues that the fuzzy logic developed by Lotfi Zadeh (1965, 1971) follows this linear pattern, and that it therefore does not capture the structure of meaning in most of natural language.

Certainly natural language can be interpreted in diverse ways. But Lakoff advances an important line of argument, and he is cited by STCM authors in justifying their approach.<sup>9</sup> Lakoff's relevance here also derives from his involvement in the early development of Zadeh's fuzzy logic and from his strong endorsement of its applications in engineering (Lakoff 2014: 11).

*Qualitative Research.* The argument that set theory captures the structure of meaning in qualitative research—an argument also evoked to justify this method—is likewise open to question. Let us summarize here a standard analytic procedure for concept formation that might be interpreted as congruent with a set-theoretic framing—and then ask if that congruence is convincing.

In qualitative analysis, and in the social sciences in general, scholars routinely seek to develop standardized, well-bounded definitions of their concepts. Further, scholars sometimes follow in their investigation the respected tradition of conceptualizing both the phenomenon and its absence in dichotomous terms. For convenience, within political science this might be called the “Sartori procedure” (Sartori 1970: 1036–40), which might appear to make a set-theoretic framing appropriate.

Scholars who take this initial step of following the Sartori procedure have perhaps three options in deciding how to use this dichotomy.

---

<sup>9</sup> Ragin (2000: 6, 171; 2008: 98); Goertz and Mahoney (2012: 16, 18); Schneider and Wagemann (2013: 21–22).

*Option 1. Dichotomies as a Commitment.* Some scholars maintain that the dichotomous conceptualization is the correct one, an argument strongly supported by Sartori and summarized by Collier and Adcock (1979). This path is followed, for example, in Przeworski et al.'s (2000) dichotomous treatment of democracy. This option would appear closely aligned with the argument that set theory captures the structure of conceptual meaning. Yet Sartori (2014), for example, specifically rejects set theory, viewing it as an unproductive analytic frame. In his own work, he introduces carefully selected tools of logic to solve very specific conceptual problems, and he strongly urges political scientists to have a knowledge of logic. But he sees the broader adoption of set theory as a turn to “technique” that can distract attention from good analysis.

*Option 2. Dichotomies as a Heuristic.* Alternatively, scholars may carry out the analysis with the dichotomous framing, viewing it simply as a heuristic. As with many dichotomies, it would be seen as a useful but false dichotomy. Here the concern about embracing set theory would be that it unproductively reifies this dichotomy.

*Option 3. Gradations.* Many researchers view the initial dichotomy as a step toward analyzing degrees and gradations. Here, the dichotomous version of STCM is not helpful, and a commonly held view might be that diverse measurement traditions other than fuzzy-sets do a better job of representing these gradations.

With all three options, there are grounds for thinking that the set-theoretic framing is inappropriate.

A final concern about the relevance of set theory is that conceptual reasoning in qualitative research—as in natural language—is routinely multidimensional. Again, this stands in contrast to the unidimensional framing of membership/nonmembership in set theory. For example, typologies are a quintessential qualitative tool, though of course also important in quantitative research (Collier, LaPorte, and Seawright 2011). A central

goal with typologies is to depict multidimensionality. In Dahl's (1971) famous typology of polyarchy, the subcategories branch out on two dimensions: lower degrees of competition and lower degrees of inclusiveness. In Linz's (1975) classic typology of authoritarianism, the subcategories are arrayed on three dimensions: participation, pluralism, and ideology/mentality. Sharply-defined ideas of membership/non-membership are sometimes useful in working with these concepts, but often they are not. Set theory might be adapted to accommodate multidimensionality, but it then the claims that that STCM is a distinctive method are further called into question.

*Case-Oriented versus Variable-Oriented; Holistic versus. Collection of Parts.* The distinction between case-oriented and variable-oriented research can readily be overdrawn. In the STCM framework, cases are analyzed in terms of dichotomous variables (i.e., conditions), and quantitative analysis sometimes relies heavily on case knowledge. In general, qualitative research does indeed address more facets of cases than quantitative analysis. However, against any plausible standard of what it might mean to analyze cases "holistically"—and not just as a collection of parts—virtually all qualitative research falls short, as does STCM.

### **3c. Concerns about Measurement**

Several concerns arise about STCM's approach to measurement.

*Strongly Anchored in Theory.* Anchoring measurement in theory is essential. However, it seems questionable to argue that by this standard, measurement in the social sciences is haphazard and unsystematic (see above). This claim might surprise several generations of measurement theorists, who have struggled over the decades to connect measurement with theory and concepts.

The type of theoretical anchoring used in set theoretic methods also raises concerns. Fuzzy-set scoring requires a well-established conception of full set membership, full non-membership, and a crossover point in between. Yet as discussed above, the ideas of membership and non-membership often poorly capture the structure of conceptual meaning. Further, if the initial designation of set membership is ambiguous and not compelling, then the rest of the scale is not convincing.

The values assigned to the variables are crucial. In the treatment of bivariate scatterplots, discussed below, it is especially important, for example, that a value of .3, or .5, or .7 be equivalent for the two variables in the scatterplot—both in conceptual terms, and in terms of measurement (Dunning 2013). Adequate justification for such equivalence is often absent, and these problems are again especially worrisome for scholars skeptical about the framing of set theory.

*Eliminating Irrelevant Variation.* In principle, the idea of eliminating irrelevant variation is valuable. However, this can come at substantial analytic and inferential cost. For example, in quantitative research on education and wealth, one periodically encounters non-linear specifications that yield insight into variability within different parts of the spectrum of values.<sup>10</sup> Such tests would not be possible if variation had been eliminated at the high or low end of these distributions. What is at stake here is perhaps not *eliminating irrelevant variation*. Rather, the fuzzy-set approach may introduce a *premature elimination of variation* that precludes these valuable tests.

*Are Fuzzy Sets Fuzzy?* The fuzzy-set scoring used in STCM is in fact not fuzzy—by the standards of Lotfi Zadeh’s fuzzy logic (Lakoff 2014: 13). Instead, it is more similar to other measurement procedures that assign scores along a linear scale. With STCM’s fuzzy-sets, the scores representing successive degrees of partial membership are given

---

<sup>10</sup> Achen (1982: 57) comments on the example of education.

fixed numerical values. By contrast, in Zadeh's approach each gradation of partial membership is treated as fuzzy. Partial membership does not have a fixed value, but rather a fuzzy range of values. STCM's fuzzy sets are not fuzzy in this sense. Rather, they correspond to more conventional scoring of variables involving the juxtaposition of two scales. The first is a dichotomy demarcated at 0.5, which separates cases that are closer to being a member than a non-member of the designated set. In addition, within the ranges of 0.0 to 0.5 and of 0.5 to 1.0, one finds equal-interval linear scales.

*Multidimensionality.* The concern that STCM does not capture multidimensionality is a measurement problem as well as a conceptual issue. Goertz and Mahoney (2012: 5, 11) evoke the "Monsieur Jourdain" idea that qualitative researchers use the language of set theory without realizing it. Instead, perhaps it could more usefully to argue that STCM researchers—in their actual practice of measurement—do not employ a one-dimensional framing that could be represented by set theory, but rather a multidimensional framing.

Multidimensionality at the level of indicators can be illustrated with five STCM studies of democracy: either as an outcome to be explained, a potential explanation, or a context of analysis.<sup>11</sup> These are serious studies within the STCM tradition, and the authors include a leading practitioner of this method. One analysis uses the Quality of Governance data compiled by Goteborg University, and two use Freedom House data. Another employs the Laakso-Taagepera index of the effective number of parties and the Gallagher disproportionality index; still one combines the World Bank voice and accountability index, Freedom House, and Vanhanen's index of voter turnout and number of parties. It would appear that the attention of these authors, like that of a great

---

<sup>11</sup> Cebotari and Vink 2013; Avdagic 2010; Berg-Schlosser 2008; Pajunen 2008; Hartmann and Kemmerzell 2010.

many scholars, is focused on diverse dimensions, and not on sharply delineated concepts that would be appropriate to set theory.

Some of these dimensions may of course be convergent. Yet as Bollen and Jackman (1989) warned some time ago, indicators of different facets of democracy can tap into very different phenomena. The point, again, is not that these are weak studies; rather, they follow the pattern of dimensional thinking that is standard in social science. A strong argument would have to be made that they are studying the same phenomenon. This argument is lacking, and the idea of clear set membership is therefore clouded.

*Scoring Procedures.* To address the challenge of establishing appropriate scores, STCM has developed procedures for the external anchoring of fuzzy-sets. Ragin illustrates calibration with the measurement of heat (2008: 72), which is standardized in relation to the freezing and boiling points of water—i.e., zero degrees and 100 degrees Centigrade.<sup>12</sup> With this example, as with additional illustrations of external anchoring, Ragin (2008: 208–212) neither makes a compelling case for external anchoring, nor convincingly differentiates it from standard reasoning about measurement in quantitative research.

*Standard Indicators.* Many STCM studies use Ragin's direct method of measurement (2008: 85), in which conventional quantitative indicators are mapped onto a fuzzy-set scale. However, too much may be lost by transforming the standard indicators, which units of measurement that are readily interpretable in substantive

---

<sup>12</sup> Given that changes in the state of water could be thought of as a dependent variable, which is centrally caused by temperature, one worries that the independent variable is being calibrated based on values of a dependent variable. This approach poses a twofold problem: first, if the analysis is focused on this same dependent variable, there is a risk of tautology; second, with a different dependent variable, this criterion may be irrelevant.

terms. Given that the fuzzy sets in fact are not fuzzy, the direct method appears virtually to replicate a scale that is linear and equal-interval. The direct method thus appears to give up too much information and to gain little. Further, it may be unnecessary, within STCM's own framework.<sup>13</sup>

*Case Knowledge and Researcher's Judgment.* Other scholars use the indirect method, in which scoring is based on case knowledge and the researcher's judgment. However, worries arise about the highly subjective character of these judgments, which require assessing precise cut-points for full membership and non-membership, as well as the cut-point at .5 and the additional steps in between. Further, for any scholar unconvinced that cases can be decisively differentiated into the categories required by set theory, the criteria for choices can be very much open to question. Simulations (see below) suggest that small differences in choices about scoring can have a substantial impact on the findings.

In sum, measurement based on the fuzzy-set method appears to be less different from standard measurement practices than is often presumed, and its distinctive features are perhaps not productive. A number of worries thus arise about measurement, and Schrodts's (2002: 453) reaction that it is the "least satisfying" aspect of STCM may be understandable.

---

<sup>13</sup> Retaining standard indicators might be justified in the framework advocated by Ragin. If STCM scholars were to retain these indicators in their analysis, then at the final step in causal inference, when—as will be explained below—the scores are dichotomized to create rows in 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.

## 4. Causal Inference

### 4a. Summary of Basic Framework

*Necessary and Sufficient Causes.* Evaluating causation is a central concern of STCM. From the beginning of this line of work, Ragin (1987) emphasized the centrality of necessary and/or sufficient causes, and this emphasis is sustained throughout this body of work—for example, Schneider and Wagemann (2012: 8)—and is pivotal in the idea of asymmetric causation discussed above.

Mahoney (2008) advocates this focus on necessary and/or sufficient—as opposed to probabilistic—causes, offering what might be called the thesis of *ex-post inevitability*. He suggests that:

the very idea of viewing causation in terms of probabilities when  $N = 1$  is problematic. At the individual case level, the ex post (objective) probability of a specific outcome occurring is either 1 or 0; that is, either the outcome will occur or it will not. To be sure, the ex ante (subjective) probability of an outcome occurring in a given case can be estimated in terms of some fraction. But the real probability of the outcome is always equal to its ex post probability, which is 1 or 0. (Mahoney 2008: 415-16)

This rejection of a probabilistic framing leads to a definition of cause for case-oriented research: “it is common to define a cause as a variable value that is necessary and/or sufficient for an outcome” (2008: 417). His argument has resonated in this literature, for example in Beach and Pedersen’s (2013: 28) claim that process tracing is inherently focused on deterministic causes; see also Ahmed and Sil (2012; 940–41).

*Causal Complexity.* Together with the emphasis on necessary and sufficient conditions, the method gives central attention to causal complexity and the idea that combinations of conditions are a key facet of complexity. Ragin emphasizes that:

Researchers who know their cases well typically understand causation conjuncturally. In fact, as a general rule, the closer analysts are to their cases, the



greater the visibility and transparency of social causation's complexity. (Ragin 2014: 84)

With regard to combinations of conditions, Ragin seeks to “assess the conjecture that there could be mixtures of four, five, or six conditions generating a qualitative change” (Marx, Rihoux and Ragin 2014: 118).

*Critique of Net Effects.* A concomitant of the analysis of causal complexity and combinations of conditions is STCM's critique of “net effects thinking” in quantitative research. The net effects approach seeks to isolate the impact of specific variables—an important goal in both conventional quantitative analysis and experimental research.

Isolating the effect of individual variables is seen by STCM as problematic and often misleading. Ragin views the attention to net effects as part of the unproductive enterprise of comparing the explanatory power of different variables and seeking to isolate their separate impact. He suggests that

the calculation of net effects dovetails with the notion that the foremost goal of social research is to assess the relative explanatory power of variables attached to competing theories. Net-effects analyses provide explicit quantitative assessment of the nonoverlapping explained variation that can be credited to each theory's variables. (2008: 178–79)

However, this approach “may create the appearance of theory adjudication in research in which such adjudication may not be necessary or even possible” (2008: 179).

The net effects focus is also seen as misleading, given that comparisons are routinely made across heterogeneous subgroups of cases. The net influence of a given variable may differ greatly across subgroups, and it is seen as far more productive to concentrate on the distinct combinations of conditions that constitute these subgroups.

Cases combine different causally relevant characteristics in different ways, and it is important to assess the consequences of these different combinations. (Ragin 2008: 181).

The net effects approach is seen as neglecting these issues.

*Truth Table.* These ideas about necessity, sufficiency, and combinations of conditions are operationalized in the truth table, the centerpiece of causal inference in STCM.<sup>14</sup> Drawn from Boolean algebra, this table is a mathematical tool conventionally used to summarize logical relationships among a series of dichotomous variables, i.e., conditions. When used in this way, the relationships it displays are logically true—hence the name.

This table might serve as a valuable data display. In contrast to a conventional data matrix, each row in the table consists not of an individual case, but rather all the cases that exhibit the same combination of binary conditions, together with the occurrence or non-occurrence of the outcome. Having the word “truth” in the name for a data display can be confusing, and Thiem and Duşa (2013, 17 n. 5) have suggested a more self-explanatory label: “table of combinations.”

In the dichotomous version of STCM, the data are already in a binary form and are directly entered into the truth table. In the fuzzy-set version, the analyst dichotomizes the fuzzy scores, often at the .5 crossover point, and enters them into the rows of the table. Based on their fuzzy scores, cases are then assigned degrees of membership in specific rows. The rows in the table are thus treated dichotomously, and gradations are retained in the form of partial membership in the rows.

The truth table helps draw attention to the three elements of causal complexity. *Asymmetric causation* is addressed by placing cases in separate rows to distinguish instances of the occurrence versus non-occurrence of the outcome. *Interactions* are displayed by juxtaposing alternative combinations of explanatory variables in distinct rows. *Equifinality* is evaluated with the multiple combinations of conditions associated with the same outcome.

---

<sup>14</sup> Zaks 2013 offers useful observations about the truth table.

Although the truth table is basically employed as a data display, rather than as a logical construct, the analyst carries out three logical operations in analyzing the table. Attention centers on: (a) *Logical remainders*<sup>15</sup>—empty rows displaying logically possible combinations of variables that are not found in a particular data set. These are understood as reflecting the “limited diversity” of the data at hand, in relation to the full combination of conditions that might potentially have been found in the data set. The number of rows increases exponentially as more explanatory conditions are added: 4 conditions yield 32 rows; 5 yield 64 rows; and 6 yield 128 rows. The majority of rows in the truth table will routinely be empty. (b) *Logical contradictions*—configurations in which the same combination of scores on the explanatory variables are associated with both the occurrence and non-occurrence of the outcome. (c) *Logical redundancies*—rows that are deemed to be subsets of other rows, and that are simplified with Boolean minimization.

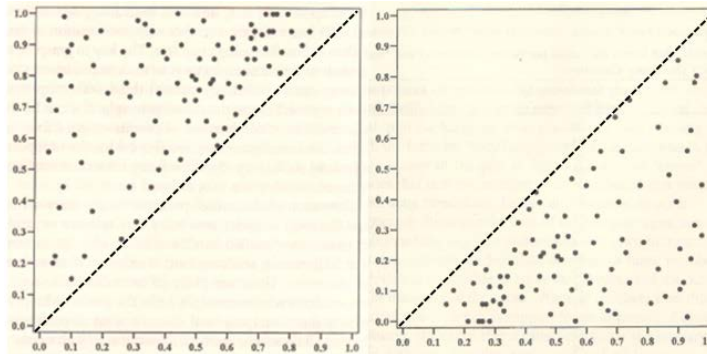
Following minimization, the truth table serves as a basis for evaluating different combinations of necessary and/or sufficient causal conditions. STCM scholars variously refer to the rows as causal paths, causal combinations, or causal recipes that capture diverse interactions among the explanatory factors associated with the outcome.

*Scatterplots: Inequalities, Gradations of Necessity and Sufficiency.* Although multivariate analysis with the truth table is the most important tool of STCM, bivariate causal inference with fuzzy sets is sometimes depicted in a two-dimensional scatterplot. Examining the treatment of the plot provides a compact way of summarizing some of the steps also taken in evaluating truth tables. Analysis of the plots centers on the two

---

<sup>15</sup> The terms for these logical relationships vary in the literature. This usage is found in Ragin 2008: 151; Rihoux and De Meur 2009: 48; and Schneider and Wagemann 2012: 106.

“off-diagonal” triangles in plots formed by the diagonal lines running from bottom left (0, 0) to top right (1,1).



**Figure 1**

**Figure 2**

**Sufficiency and Necessity Tests Based on Scatterplots**

Source: Ragin 2000: 215, 236.

The evaluation of causal conditions is formulated in terms of inequalities. Sufficiency is demonstrated if the data are consistently located in the upper-left triangle—i.e.,  $y$  is always greater than or equal to  $x$  (Figure 1). Necessity is demonstrated if the data are located in the lower-right triangle in the plot—i.e.,  $x$  is always greater than or equal to  $y$  (Figure 2). This interpretation relies on two key ideas of fuzzy-set analysis: (1) *measurement equivalence*, i.e., the understanding that equal scores on  $x$  and  $y$  denote equivalence in substantive terms—in this case, equal degrees of closeness to full set membership in the two phenomena being measured; and (2) *subset relations*, which are fundamental to fuzzy logic and provide the rationale for this framing of sufficiency and necessity; and (2) (Ragin 2000: 214–18).

Important refinements, involving the ideas of “usually” and “almost always” necessary or sufficient, along with “probabilistically” necessary or sufficient, have been added to the analysis of the scatterplots—and more broadly are central to causal

inference in STCM.<sup>16</sup> In analyzing sufficiency, for example, the researcher calculates the proportion of the cases that fall outside of the upper-left triangle. The significance of this proportion is tested against the benchmark proportion established by the investigator—for example, 0.65 might be treated as usually necessary and 0.05 as the “significance” level (Ragin 2000: 227-229). Cases might be located outside the triangle due to “imperfect evidence,” i.e., “error, chance, randomness, and other factors” (Ragin 2000: 109), and these criteria provide some protection against false negatives. More recently, Ragin (2008: 45-52) has formulated tests based on consistency scores, which likewise assess the degree to which this off-diagonal pattern is followed. A perfect consistency score would be 1.0; a strong score 0.8; and a low score 0.2. If the consistency score for a test of sufficiency (as in Figure 1) is high, yet lower than 1.0, the causal condition may be designated as almost always sufficient (Ragin 2008: 49).

*Qualitative versus Quantitative.* Based on these arguments concerning sets and logic, STCM scholars draw a sharp distinction between qualitative and quantitative approaches. Qualitative work is viewed as based on logic and set theory; quantitative methods are based on probability theory. The distinction is central to Ragin’s approach and is emphasized by Goertz and Mahoney (2012, chap. 2).

These contrasts between qualitative and quantitative methods are closely connected with three further distinctions already discussed in Part 3: case-oriented versus variable-oriented research, kinds of cases versus relationships among variables, and the holistic versus “collection of parts” view of cases (Ragin 1987, 2000, 2008: *passim*).

---

<sup>16</sup> The usually and almost always forms are used periodically in Ragin 2000 and 2008. Ragin actually uses the expression “probabilistic assessment” of necessity and sufficiency; Mahoney (2008) speaks of “probabilistically” necessary/sufficient. See also Mahoney (2010: 135; 2012: 25, n. 10)

*Algorithms.* STCM employs a fairly elaborate range of analytic procedures i.e., algorithms, only some of which are discussed here. A number of elements have been added since Ragin's initial formulation in 1987, such that what began as a relatively simple method has become far more complex. Given that reliance on elaborate algorithms has become a point of concern in new debates on causal inference, this point merits emphasis.

*Innovative Work on Process Tracing.* Finally, the most innovative area of new work on STCM is the effort to strengthen the key step from "association to causation," based on case studies and process tracing. This work—most compellingly that of Schneider and Rohlfing<sup>17</sup>—seeks to open the "black box" of the truth table with close analysis of causal connections and mechanisms. Especially given the pervasive emphasis in STCM on assessing causal claims, this step in getting from overall patterns in the data to strongly-grounded causal assessment is crucial.

Schneider and Rohlfing propose criteria for selecting the cases for process tracing that will be high yield within the STCM framework. To evaluate and improve theory, they recommend looking at both typical and deviant cases. At a more fine-grained level, they offer several principles for case selection: maximum set membership, maximum set membership difference, max-max difference, and maxi-min difference. Their approach is thus carefully articulated with analytic procedures of STCM.

Their initiative thus closely parallels the earlier efforts of Gerring (2008) and Seawright and Gerring (2008), who use case studies to increase inferential leverage in conventional quantitative research. These two scholars likewise map out criteria for selecting cases that will be high yield. They suggest looking at typical, extreme, deviant,

---

<sup>17</sup> Schneider and Rohlfing 2013; Rohlfing and Schneider 2013.

or influential cases, based on the position of the case in relation to an initial set of medium- to large-N findings. For both methodological traditions, these are valuable innovations.

These parallel efforts in the two areas of methodology merit close attention and further development.

#### **4b. Concerns about Causal Inference**

A number of misgivings arise about causal inference. This section first considers a broad issue—whether the findings of STCM are interesting—and then addresses other questions, including the justification for set theory, the truth table, and simulation findings.

##### ***Are the Findings Interesting?***

Are STCM findings substantively interesting? In fact, they may often not offer the kinds of insights many researchers seek. The discussion of conceptualization and measurement in Parts 3b and 3c has already explored some of these problems: real-world units of measurement are not employed; information is lost due to dichotomization—even in the fuzzy set version; and the substantive equivalence of scores across different indicators is inadequately justified. For causal assessment as well, the substantive pay-off too often seems questionable.

*Does “Causal Complexity” Become Too Complex?* Ideas of causal complexity are inherently intriguing for social scientists. However, STCM sometimes addresses too many factors, lending credence to Schrodts’s (2002: 453) observation that STCM findings sometimes exhibit “mind-numbing intricacy.”

Two issues arise here. First, it is valuable to look for complexity, yet also essential to have tests in which one can fail to find it. Although causal patterns definitely are sometimes complex—possibly encompassing the interaction among as many as six or more explanatory factors—at times they may not be. Tests are needed to discern the difference. In this regard, STCM’s apparent tendency to generate false positives—suggested by the simulation findings discussed below—is a matter of concern.

Second, scholars routinely have a substantive interest in nailing down the interaction not among numerous variables, but among two or three. STCM researchers have advocated this method for public policy research, yet policy researchers sometimes wish to analyze the interaction among only a small number of factors. For example, they may be interested in contextual effects involving contrasts in a given two-variable relationship as it is manifested in different policy settings (Tanner 2014). Analysis focused on more limited degree of complexity is highly salient for social science researchers in general.

*Rejection of Net Effects and Inattention to Relative Importance of Variables.* Many scholars seek to assess the effect of individual variables, and in an imperfect world of data analysis they tease them out as best they can. Further, compared to the daunting challenge of analyzing the interaction among as many as six variables, the (admittedly imperfect) option of concentrating on the net effect of single variables is a valuable alternative.

STCM’s rejection of net effects stems in part from a mischaracterization of standard practices in conventional quantitative research. Ragin treats the net effects approach as part of the larger enterprise of assessing the relative explanatory power of the variables included in a regression analysis. This assessment was traditionally based



on comparison of *standardized* regression coefficients, along with assessment of the contribution of different variables to the overall variance explained.

Yet for political science, at least since Achen's 1977 article "Perils of the Correlation Coefficient," the *unstandardized* coefficient is the preferred option. The unstandardized coefficient specifically *does not* lend itself to comparison of the explanatory contribution of several variables to the overall  $R^2$  (or adjusted  $R^2$ ). Evaluating  $R^2$ , in turn, is now seen to have questionable value.

By contrast, the unstandardized coefficient *does* take a step toward assessing something that is much more interesting in substantive terms: the effect of a given explanatory variable, expressed in terms of the real, unstandardized units of measurement employed for each variable. It is about "real-world relationships." If the researcher proceeds with caution, valuable comparisons of different variables' importance can be achieved.

The STCM critique is of course correct in suggesting that a net effect can be meaningless if estimated across heterogeneous subgroups. But a long tradition in political science and sociology has given careful attention to contexts—which can easily involve "contexts within contexts." Przeworski and Teune's (1970: 43–46) treatment of "comparing relationships" was an early statement in this tradition, though hardly the first. R. Collier (1982: 76–94) provides an illustration of decomposing coefficients by context; and the wider literature on contexts and contextual effects is discussed in Part 9 below.

As Achen (2014: 26) has emphasized, estimates of effects can be distorted if context is not adequately taken into account, and one can readily make mistakes. However, these challenges are no more daunting than a great many of those faced by STCM, and the rejection of net effects appears to abandon an analytic focus of greater substantive interest than the standard findings of STCM.

Relatedly, researchers often want to know whether, in a given causal process, some variables are more important than others. In the critique of net effects, STCM rejects the idea that this is valuable information. Yet in an analysis that finds four, five, or six variables to be relevant, it is difficult to imagine that some are not more influential than others in shaping the outcome. In the tradition of regression analysis, evaluating relative importance is certainly more difficult than has sometimes been recognized; however tools do exist for addressing this issue.

*Diluting the Ideas of Necessity and Sufficiency.* The ideas of necessity and sufficiency—jointly with the idea of asymmetric causation—are inherently intriguing to many scholars. Yet the treatment of these patterns in terms of inequalities (see discussion of scatterplots above) may dilute them to the point that that their substantive appeal to non-STCM scholars can be lost. In the scatterplots above, necessity is established if the score on the explanation is always greater than or equal to the score on the outcome. This criterion builds directly on the understanding of subset relations that is fundamental to Boolean logic. In STCM’s own framework, these ideas are appropriate.

However, this analytic procedure may well take scholars away from what is inherently intriguing about these hypothesized causal patterns. For example, by no bourgeoisie, no democracy, did Barrington Moore mean that the fuzzy-set score for “bourgeoisie” is consistently (or at least, almost consistently) equal to or greater than the fuzzy-set score on democracy? This formulation is extremely remote from the “aha” feeling produced for many scholars by Barrington Moore’s terse formulation of this relationship.

The same problems can arise with the ideas of almost always necessary or sufficient, or probabilistically necessary or sufficient. These conceptions may well not

capture the core ideas that made necessity and sufficiency appealing ideas to begin with.

These practices might be justified because they solve another problem. They help avoid false negatives, which can arise if the researcher is analyzing data containing error. On the other hand, these procedures do not protect against false positives, leaving STCM open to confirmation bias. The justification for weakening the substantive interest of findings—with goal of addressing other methodological priorities—is thus partly undermined.

### ***Justification for Set Theory: The Ex-Post Inevitability Thesis***

The justification for embracing set theory as an approach to conceptualization and measurement has already been questioned in Parts 3b and 3c. Concern also arises about the set-theoretic framing of causal inference.

The *ex-post inevitability* thesis discussed above maintains that once an outcome has occurred in case-based research, the realized probability of the outcome is 1.0. In retrospect, it is inevitable. This thesis maintains that a deterministic view of causation is therefore appropriate, centered on necessary and/or sufficient conditions. A probabilistic view is inappropriate.

This approach neglects the challenge of inferring the underlying causal process. At a given point in time, the “real world” definitely yields only one outcome per case, but that does not demonstrate that the underlying causal relationship is deterministic.

To be clear, the issue here does not hinge on the distinction between explaining the outcome in a given case, versus inferring an explanation intended to have broader generality. Even in case-based work, one must make inferences about the underlying process that generates the case-specific outcome.

Regarding the ex-post inevitability thesis, what if the outcome of interest has a numerical value, rather than being in a dichotomy? For example, the researcher may ask why a country has an annual per capita GNP of \$20,000. Is it interesting, productive, or even plausible to argue that the constellation of conditions found in that country is necessary and/or sufficient to explain the \$20,000—as opposed to the outcome of \$20,100, or \$20,010, or even \$20,001?

To underscore a final point: One does not have to be, in any technical sense, a “probabilist” to recognize that necessity and sufficiency may not offer a useful perspective on these issues. Alker’s (1973: 307) engaging claim that “actualities are low probability events” needs to be supplemented with the observation that probabilities may also be high, or intermediate. But centering attention on probabilities provides a more fruitful line of discussion than insisting on necessity and sufficiency.

### ***Sharp Distinction between Qualitative and Quantitative***

More broadly, the sharply drawn distinction between qualitative and quantitative analysis makes this distinction too rigid. It also diverges from the basic perspective that animates multi-method research.<sup>18</sup> From a multi-method perspective, the synergistic relationship between qualitative and quantitative analysis is a key area of methodological innovation. This idea is forcefully expressed in the subtitle of Ridenour and Newman’s (2008) book: *Exploring the Interactive Continuum*. Certainly, qualitative and quantitative analysis can contribute different forms of analytic leverage. But emphasizing a sharp division between these approaches seems inappropriate and unproductive.

---

<sup>18</sup> For example, Lieberman 2003 and 2005; Brady and Collier 2004/2010; Gerring 2008; Seawright 2015.

Questions likewise arise about the strong insistence on the contrasting foundations of the two methods: set theory and logic in qualitative methods, and probability theory in quantitative methods. In fact, probability theory is based on set theory, and formal logic is frequently used by quantitative researchers. The Empirical Implications of Theoretical Models movement<sup>19</sup> (EITM) integrates the set-theoretic framing of game theory with the probabilistic tools of quantitative analysis, and a parallel integration appears in the macroeconomic literature. Thus, both the qualitative *and* the quantitative traditions have an important foundation in logic. The difference resides not in whether they have this foundation, but in what they do with it.

This argument for a sharp division between qualitative and quantitative also overlooks the fact that probabilistic ideas and the idea of partial causes are important in qualitative work. Goertz's (2003) valuable inventory of necessary condition hypotheses productively challenges methodologists to look closely at the causal language used by researchers. Let us follow his example and consider an illustration: Tannenwald's (1999) article on the "Nuclear Taboo," a study often cited as an excellent example of process tracing. She analyzes the horrified reaction of top U.S. policy-makers to the United States' use of nuclear weapons in World War II. This reaction is hypothesized to have generated a nuclear taboo that strongly influenced subsequent decisions not to use nuclear weapons.

In Tannenwald's analysis, most of the causal language used expresses the idea that given factors will *decrease* or *increase* the degree to which an outcome will occur, involving incremental effects. Deterministic language is rare. Examples of decreasing are *constrain* (21 instances), *inhibit* (11), and *limit* (3); examples of increasing are *encourage* (2), *raise* (2), and *bolster* (1). This imbalance between decreasing and

---

<sup>19</sup> Granato and Scioli (2004).

increasing makes sense in substantive terms, given the argument that the nuclear taboo decreased the likelihood that nuclear weapons would be used. The idea of asymmetric causation is thereby captured, and without the detour of discussing necessity and sufficiency.

Some terms directly express a probabilistic idea: *likely* (5), *probability* (2), and *unlikely* (2). Few terms express causal necessity or sufficiency. The only two examples found in this search are *contribute decisively to* (1) and *prevent* (1). Further, in another instance deterministic language is explicitly rejected in favor of more incremental phrasing. Thus, “norms do not determine outcomes, they shape the realm of possibility.”<sup>20</sup>

Ideas of partial and probabilistic causation are certainly difficult to summarize with precision in qualitative studies, but this does not mean that they are not important. It is indeed a weakness of qualitative research—as Collier and Collier (1991: 20) note—that “it lacks a precise means of summarizing relationships in terms that are probabilistic rather than deterministic.” The researcher must therefore “rely on historical analysis and common sense,” thereby “recognizing that the relationships under analysis are probabilistic and partial.” While formal tools for assessing probabilities are lacking, these ideas are important for qualitative work.

### ***Problematic Implications of Truth Table as a Logical Construct***

As noted, in the STCM framework three concomitants of working with the truth table are the analysis of logical remainders, logical contradictions, and logical redundancies. For scholars outside the STCM tradition, the treatment of these logical patterns as the basis for empirical analysis can appear counterintuitive.

---

<sup>20</sup> Documentation of this analysis of Tannenwald, as well as of other substantive studies in the qualitative tradition that are often cited in this methodological literature, will be posted online.

Consider the first pattern—the logical remainders, i.e., empty rows in the table, which to reiterate increase exponentially with more explanatory conditions. The empty rows call, in principle, for counterfactual reasoning about the often numerous combinations of conditions not found in the cases being analyzed. QCA software offers a variety of options for dealing with the empty rows, including filling them in with hypothetical values. Counterfactual reasoning is crucial in contemporary thinking about causal inference. However, this kind of counterfactual reasoning does not correspond to that called for by the potential outcomes framework for causal inference, discussed below, that has emerged in the past three decades.

This concern with empty rows is also sharply divergent, for example, from standard norms in the field of comparative-historical analysis, for which STCM is intended to have great value (Ragin 1987; Marx, Rihoux, and Ragin 2014). For example, in their book *Shaping the Political Arena*, R. Collier and D. Collier (1991) would not possibly have wanted to—or been able to—address the equivalent of a large number of “empty rows” that would have arisen by considering all combinations of the explanatory variables employed in the analysis. Certainly some books have a concluding chapter that places the cases analyzed in a wider comparative framework—including potentially some commentary on what are in effect 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, and it would seem implausible that it should be.

If the truth table were simply treated as a valuable data display, perhaps the analysis of empty rows could be dropped.

The treatment of logical redundancies also merits comment. Serious problems arise about the Quine-McCluskey algorithm, which eliminates logical redundancies by

combining rows in the truth table that are subsets of other rows. A key issue with Quine-McCluskey is sensitivity to error in data. Whereas error is far less an issue in engineering applications, for which Quine-McCluskey was originally designed, it is a major issue in social science applications.

### ***Are Causal Inferences Unstable and Prone to Error?***

Simulation studies raise serious questions about STCM.<sup>21</sup> These studies report great sensitivity to measurement error, and findings are also unstable in response to small differences in the parameters that must be set in applying STCM's algorithms. With a known or simulated data-generating process (DGP), the method appears highly vulnerable to mistaken inferences, including false positives. It often fails in capturing causal complexity in the DGP, even when this complexity is structured in a way that corresponds to the patterns STCM is designed to detect (Krogslund and Michel 2014).

Schneider and Wagemann discuss simulations and robustness in a more encouraging light, but they conclude: "QCA is not vastly inferior to other comparative methods in the social sciences" (2012:294). Given sharp criticism noted below of conventional quantitative analysis, which is a key method of comparison, this is faint praise.

How should scholars assess these simulations? First, they should consider whether the simulations are appropriate to STCM. Simulations must fit the method being evaluated, and scholars should and will scrutinize this fit.<sup>22</sup> In this respect, the new simulations are a major step forward. They use STCM software and carefully seek to match the simulation to analytic procedures of the set-theoretic approach.

---

<sup>21</sup> Hug 2013; Krogslund, Choi, and Poertner 2015; Krogslund and Michel 2014; Kurtz 2013; Lucas and Szatrowski 2014; Seawright 2013.

<sup>22</sup> For example, Ragin and Rihoux (2004); Ragin (2014); Thiem (2014).



Unquestionably, there is room for improvement and refinement in future simulations, yet the cumulative evidence of unstable findings and error raises major concerns.

Second, scholars must ask if one might expect that STCM findings will be unstable and prone to error. False positives are a central concern, and the discussion of scatterplots above makes it clear why this might be a problem. STCM is specifically designed to address false negatives. The adjectives “usually,” “almost always,” and “probabilistically” are attached to necessity and sufficiency to accommodate the possibility that the true, underlying relationship of necessity and sufficiency was not fully revealed in the observed, imperfect data.

No corresponding procedure is offered to address false positives. For example, in the scatterplots discussed above, some or many cases might be found in one of the two triangles—again, due to imperfect evidence, error, and randomness—even when the underlying relationship is not one of necessity and sufficiency. What scholars need is a probabilistic approach to guard against false positives. This would inevitably, and perhaps fruitfully, increase the convergence between STCM and conventional quantitative methods.

Regarding instability of findings, another concern is dichotomization. Even with the fuzzy-set version that incorporates gradations, the algorithms ultimately reduce the rows in the truth table to dichotomies, which can be more vulnerable to error than gradations (Elkins 2000:298–99). The number of cases per causal path is often surprisingly small, pointing to questions about the stability of findings.

Unstable findings might also be expected, given substantial evidence that the basic STCM algorithms are more similar to those of conventional quantitative methods than has been recognized (e.g. Paine 2015; Seawright 2005). When conventional methods are applied to a small or medium N, one expects unstable findings, so it might not be

surprising that this also occurs with STCM. In a sense, everyone is “in the same boat,” but great caution is needed in reporting findings—a caution that does not appear to be evident in, or built into, STCM.

### ***Current Standards for Causal Inference***

These problematic simulation findings should also be seen in light of the wide-ranging discussion of standards for causal inference that has emerged over the past three decades. This challenge—which has coincided with the emergence and development of STCM—is relevant to diverse methods.

These new standards are anchored in what is often called the Neyman-Rubin-Holland model,<sup>23</sup> sometimes referred to as the “potential outcomes” framework. This approach posits a fundamental problem of causal inference, involving a specific definition of a causal effect as the difference between (a) the outcome observed in a particular case in which a given causal factor was present, and (b) the outcome that would have occurred in that same case had the causal factor not been present. This counterfactual idea calls for careful reasoning about what the case would have been like, had the posited cause not been present. This could involve dichotomous alternatives, but is not restricted to them.

This approach is accompanied by a series of research priorities. These include a new caution about inferences from observational data, and indeed from all types of data, along with a commitment to simple analytic procedures and parsimony and skepticism about complex algorithms. Achen’s (2002:446) “rule of three” mandates parsimony in regression analysis: it is “meaningless” to test more than three explanatory

---

<sup>23</sup> Brady 2008; Sekhon 2008; also Goertz and Mahoney (2012).

factors. Some feel that three may be too restrictive, but as a broad guideline this mandate is definitely valuable, including for STCM.

Further, STCM relies on precisely the kind of complex algorithms that have been the focus of concern in new thinking on causal inference. This problem has been exacerbated by the widespread availability of STCM software that allows scholars to apply the algorithms with relatively little reliance on case knowledge. Indeed, in the trajectory of other innovative methods, the widespread availability of software has been 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—vis-à-vis what was originally mapped out by Ragin—has occurred with STCM. This is of course a comment on how STCM is practiced, and not on the underlying principles of the method. However, when one sees a listing of 383 substantive studies using this method, there are certainly grounds for asking what the STCM research program is accomplishing.

### ***Adding Process-Tracing: Limited to Necessity/Sufficiency?***

Viewed in light of these standards for causal inference, Schneider and Rohlfing's new work on process tracing is a crucial effort to strengthen STCM. Their goal is to carry out in-depth evaluation of alternative causal patterns to achieve greater insight into the findings of necessity/sufficiency.

A key question arises here: is necessity and/or sufficiency tested, or is this finding built into the analysis? The contribution of process-tracing tests to causal inference has more leverage if, in principle, it is possible to conclude that the relationship is causal, but not one of necessity/sufficiency. Against that alternative, if the test demonstrates that necessity/sufficiency is found, the test is all the more powerful.

However, an important theme in this literature is that causal findings from process-tracing tests—and from case-study analysis in general—inherently entail necessity/sufficiency (Mahoney 2012: 573; Beach and Petersen 2013: 28; Blatter and Haverland 2014: 9). Schneider and Rohlfing (2013: 569; also Rohlfing and Schneider 2014) appear to adopt the same position.

If process tracing only yields findings of necessity/sufficiency, then the contribution of these tests within the Schneider-Rohlfing framework is greatly reduced. The position taken here is that (a) the results of process-tracing tests are not limited to necessity/sufficiency; and (b) much work is needed to find criteria for analyzing causal processes and mechanisms, with the goal of differentiating among causal connections that reflect probabilistic versus necessary/sufficient patterns.

### **5. Methodological Path Dependence?**

To place STCM in perspective, it is productive to look back at the context in which Ragin first formulated his critique of conventional quantitative methods and the proposed alternatives (Ragin 1987; also Ragin, Mayer, and Drass 1984). At that time, certain alternative tools were available which appeared plausible; yet subsequently other tools have emerged—or been refined—that may offer better solutions to some of the same analytic problems. What has occurred may well be a kind of methodological path dependence, involving the persistence of tools that have subsequently been superseded by better alternatives.

#### ***Evolving Protests***

Ragin's critique of quantitative methods echoed a wider protest against methodological narrowness that had begun earlier. Parallel critiques had already emerged in the 1970s,

reflected for example in criticism of the narrow conception of variables. In the early 1970s, the disparaging label “1960s variable analyst” was sometimes applied informally by mainstream methodologists to more “old fashioned” scholars who engaged in narrow, correlation-based, atheoretical, quantitative research.<sup>24</sup>

A prominent voice in this period—beginning with a 1977 article on “The Perils of the Correlation Coefficient”—is the highly influential methodologist, Christopher Achen.<sup>25</sup> Recurring themes in Achen’s writing are nearly identical to those enumerated above as the central critique advanced by STCM scholars. Thus, conventional methods are seen by Achen as: (1) naïvely variable-oriented and insufficiently case-oriented; (2) relying on an inadequate understanding of variance; and (3) neglecting causal complexity, given the excessive reliance on a linear-additive model and the failure to incorporate contextual effects, interactions, and other interdependencies among explanatory factors; (4) failing to gain leverage in addressing these limitations by building on an iterated examination of theory and case knowledge.<sup>26</sup> The critical role of substantive knowledge has been a fundamental theme throughout Achen’s writing—expressed yet again in his recent observation that “ it is hard to learn much about the value of a proposed new estimator when the substantive model under test is brutalizing the data” (Achen 2014: 26).

Somewhat later, around the same time as Ragin, quite a few other authors developed similar arguments. Examples include Lieberson’s (1987) wide-ranging

---

<sup>24</sup> Use of this disparaging label at that time was observed by David Collier.

<sup>25</sup> Achen was, for example, the first president of the APSA Organized Section for Political Methodology.

<sup>26</sup> Achen 1977: 807, 812–13, 814; 1982: 7, 12, 66, 67, 69, 70; 1983: 87. See also 1992: 198, 206, 207, 209.

critique of the naïve understanding of variance and Abbott's (1988) challenge to the linear-additive model. These critiques had become familiar and standard to the point that in 1992, Achen commented with reference to Ragin's 1987 book that "Ragin's arguments are very familiar, indeed nearly clichés" (1992: 197).

### ***Path Dependence***

The point here is definitely *not* that by the 1980s, the critiques of conventional quantitative methods were no longer necessary—indeed, conventional quantitative methods have an inertial momentum that persists today. Rather, the issue is the methodological alternatives available to Ragin at the time he formulated his critique—as opposed to the range of options that have become available subsequently.

In the 1980s, Ragin actively searched for alternatives to standard quantitative approaches. He responded to the limitations of these approaches by introducing Boolean methods, based on logic and set theory, which are "designed to overcome these shortcomings" (Ragin, Mayer, and Drass 1984: 222). He thus sought "a synthetic, broadly comparative strategy" that would meet a twofold requirement: it "must be both holistic—so that the cases themselves are not lost in the research process—and analytic—so that more than a few cases can be comprehended and modest generalization is possible" (Ragin 1987: xiv).

In conjunction with the adoption of Boolean methods, Ragin turned to truth tables and the Quine-McCluskey algorithm—which originated in electrical engineering—as the basic tool for the logical minimization that is a central step in working with truth tables.

Somewhat later, in response to the challenge that set theory was limited by the use of dichotomous variables, Ragin (2000) turned to the tradition of fuzzy-set analysis that had originated with Lotfi Zadeh's pioneering work on fuzzy sets. This approach, also

created in the field of electrical engineering, offered algorithms anchored in set theory, but which allowed for gradations in set membership. Multi-valued STCM (Cronqvist and Berg-Schlosser 2009) has also been developed with the goal of moving beyond the original dichotomous version, but fuzzy-sets are the major tool now employed for analyzing graded membership.

The truth table, Quine McCluskey, and fuzzy sets may well have been plausible options at the time they were adopted. Yet better alternatives are available today. For the many scholars who embrace STCM's overall analytic goals, the persistence of these earlier tools appears problematic.

## **6. Better Options**

Promising alternatives are available for addressing the analytic challenges that motivated the development of STCM. This section considers three groups of tools that respond to these challenges, and that have evolved substantially since STCM made its basic choices about analytic procedures: (1) Traditional methods that have recently seen substantial innovation: process-tracing, contextualized comparison, cross-tabulations, and natural experiments. (2) Algorithmic tools, which in contrast to the first group rely on complex statistical procedures (algorithms): Classification and Regression Trees (CART), new approaches to interactions, and logistic regression. (3) The analysis of mechanisms.

Three of the tools in the first group might be seen as “tried and true” methods—which recently have seen substantial innovation. Natural experiments are considerably newer, but are by now are also a well-established method. Those in the second group are more problematic and are presented here with many caveats. We seek to be candid about their limitations, just as scholars need to be candid about the limitations of STCM.

Finally, the discussion of mechanisms brings together issues and challenges faced by both (1) and (2), as well as by STCM.

### **6a. Innovation in Traditional Tools**

Both Seawright (2005: 24) and Lucas and Szatrowski (2014: 2) have suggested that scholars should consider turning from STCM to more traditional tools of qualitative analysis. Several traditional tools tackle the same challenges addressed by set-theoretic methods—perhaps with greater success in some respects. These tools generally focus on only one or two these challenges, which might be seen as a limitation of the alternative methods. On the other hand, STCM may try to accomplish too many things at once, and doing a more adequate job in addressing specific challenges may be more productive. Especially when joined with a strong research design, more traditional qualitative methods can provide a stronger basis for achieving the analytic goals of STCM.

*Process Tracing.* New work on process tracing is perhaps the most valuable area of innovation in the field of qualitative and multi-method research. Humphreys and Jacobs (2013), along with Bennett (2014), have taken the general idea of a Bayesian approach and have made it more specific and useful. In notable contrast, STCM has taken what is seen here as the unproductive initiative of treating causal patterns in process tracing as inherently involving necessary and sufficient causes.

Humphreys, Jacobs, and Bennett offer new criteria for evaluating process-tracing evidence, based on Bayesian analysis. They situate the four traditional process-tracing procedures—straw-in-the-wind, hoop, smoking-gun, and doubly-decisive tests (Van Evera 1994)—on a spectrum of alternatives, based on the degree to which: (a) finding or not finding evidence that affirms a hypothesis has (b) a modest or strong effect on (c)



the posterior assessment of whether the hypothesis is supported. For many scholars who have worked with process tracing, this innovation is a breath of fresh air. This framework moves the discussion beyond categorical divisions among the four tests. It also supersedes the earlier characterization of the tests as providing necessary and/or sufficient criteria for affirming a causal inference. In the new framework, that characterization is no longer useful.

A central question is whether researchers should “fill in the numbers,” using the Bayesian algorithms to actually make calculations. Instead, they can employ the method as a useful heuristic. Filling in the numbers necessitates modeling decisions by requiring the researcher’s opinions (such as priors) to be represented as formal probability distributions. To us, the simple heuristic option appears preferable.

*Contextualized Measurement.* A long methodological tradition has explored the interplay between context and measurement in a way that is more compelling than the measurement procedures and treatment of context in STCM. Comparability is seen as dependent on the interplay between the concepts employed and the contexts compared (Przewroski and Teune 1970; Sartori 1970). Examples from the comparative case-study tradition include Locke and Thelen’s (1995) comparison of trade-union responses to economic flexibilization in advanced industrial countries; and Skocpol’s (1995) demonstration that the U.S. has been mischaracterized as a “welfare laggard,” due to an inappropriate choice of the domain of observation used in measuring timing. This tradition of comparison is has been advanced by van Deth’s (1998/2013) book *The Problem of Equivalence*, which was reissued in the “Classics Series” of the European Consortium for Political Research Press.

Since 2000, a wide-ranging literature has emerged on the comparability and “harmonization” of a broad spectrum of economic, social, and political indicators. An

important part of this innovative work is carried out by international organizations, an excellent example being a World Bank project on measuring poverty in Latin America and the Caribbean (World Bank 2014).<sup>27</sup>

In contrast to the STCM tradition, these studies since 2000 generally seek to retain real-world units of measurement. They are based on complex, cross-national data sets and raise issues very much parallel to—yet somewhat distinct from—the treatment of comparability in the comparative case-study tradition. Adcock and Collier (2001) took a step toward integrating these two lines of work and connecting them with a shared framework for evaluating measurement validity in both qualitative and quantitative research. A caveat here: one key idea in the general literature is that “validity” is not a readily achievable end state; rather, scholars engage in an ongoing task of “validation.” This will continue to be a fruitful area of ongoing methodological innovation as scholars address this and other challenges.

*Cross-Tabulations: Traditional Tools, New Leverage.* Leading methodologists—Achen (2002:2) and Freedman (2008:241)—have advocated cross-tabulations as a methodological tool. Far from being old-fashioned, these tables are seen as valuable and revealing, and they can be connected with experimental designs that yield strong inferential power.

Innovative work with cross-tabulations, that includes close attention to interactions and context, extends back over many decades (Lazarsfeld 1955). Ongoing work continues to illustrate insightful and refreshingly transparent analysis of interactions—for example, Mauldon et al.’s (2000) evaluation of two alternative welfare interventions, aimed helping teenaged mothers graduate from high school. Based on a

---

<sup>27</sup> World Bank. 2014. *Social Gains in the Balance: A Fiscal Policy Challenge for Latin America and the Caribbean*. Washington, D.C.: World Bank Document 85162.

simple 2x2 cross-tabulation, these authors find that one of these interventions by itself had no value, the other by itself had some value; and the two of them together nearly doubled the high-school graduation rate, compared to the absence of either intervention (2000: 35). They thereby demonstrate an interaction between two variables that some readers might see as more substantively interesting than the efforts, noted above, to analyze the interactions among as many as six variables—or even more.

This example also points to recent gains in simplicity and credibility that stem from strong research design: Mauldon et al.'s (2000) randomized experiment obviates standard concerns about confounding and causal order.

*Case Knowledge and Natural Experiments.* Intensive use of case knowledge is a basic theme in most of the literatures discussed in this article. Natural experiments illustrate how case knowledge can be linked with a method of causal inference that meets the standards discussed in Part 4b above (Dunning 2012). The basic set-up for natural experiments is to identify real-world situations of “as-if” random assignment of cases to “treatment”—i.e., to alternative values of an hypothesized explanatory variable. Designation of a given process as as-if random depends heavily on strong case knowledge, and this joining of qualitative evidence and rigorous inference could readily be seen as more compelling than the much discussed use of case knowledge in STCM.

Natural experiments also have the merit of mandating extremely simple data analysis and data displays—for example, comparison of frequencies and percentages. Further, they present valuable opportunities to study institutions and macro-processes—in contrast to many true experiments, which tend to concentrate on individual behavior.

Actual studies vary greatly in their success in meeting three criteria of evaluation: the plausibility of as-if random assignment, the credibility of the statistical model, and the real-world relevance of the causal pattern being evaluated (Dunning 2012). Failures

as well as successes are evident in the literature. Pursuing a natural experiment in many circumstances can provide a good answer to challenges of research design, but this design can readily fail in meeting the three criteria.

## **6b. Algorithmic-Based Tools**

This second group of methods—based on elaborate statistical procedures—addresses some or all of the broad analytic goals advocated by STCM. As with STCM, the goal here is best understood as the descriptive search for patterns in the data. Analysts must leverage additional information on causal process and use stronger research designs to move toward any kind of plausible claim of uncovering causal relationships.

*Classification and Regression Trees.* One alternative to the analytic procedures used by STCM are Classification and Regression Trees (CART). Breiman et al. introduced this method in 1984, at approximately the same time that STCM scholars began to utilize the truth table and the Quine-McCluskey algorithm. Random Forests (RF), subsequently introduced by Breiman in 2001, is a valuable extension of CART. This further elaboration uses resampling techniques to carry out numerous iterations of the analysis, based on random subsets of the data.

These methods are relevant because they have capabilities of interest to STCM scholars: analyzing complex data, nonlinear relationships, higher-order interactions, and divergent patterns in different contexts. Further, they appear to do a much better job than STCM of analyzing these patterns in the data. Seawright (2013) finds that CART introduces much less error in data analysis than STCM. Kroglund and Michel (2014) have tested these methods based on a simulated data set which contains the structure of interest to STCM scholars. Treating this data set as representing a known “data generating process” (DGP), they find that— for the small to medium N of interest to

STCM—Random Forests greatly increases the validity of findings in recovering this underlying DGP.

While CART should definitely not be oversold, it addresses shortcomings of STCM that have been discussed throughout this article--especially those deriving from truth tables and the Quine-McCluskey algorithm. (1) With regard to measurement, CART can accommodate data that is nominal, ordinal or continuous and does not require the dichotomization of variables. (2) It can test multiple, alternative cut points in a given variable and can also show that different cut points are relevant for gaining insight into different patterns of interaction. It can thus sidestep the “premature elimination” of variance mentioned above, resulting from dichotomization within the framework of set theory. (3) In comparison with the truth table, it can readily analyze a larger number of independent variables. (4) Findings based on CART are readily interpretable in substantive terms and are easy to grasp. (5) Interactions can be understood as causal paths if the researcher so desires, but other forms of interactions can also be analyzed. (6) CART differentiates between variables that are more important and less important—based on a measure of the proportional reduction in error at each step in the analysis.

To reiterate a key caveat: as with STCM and regression analysis, CART yields only associations within a given data set. Like SCTM, the technique does not solve many of the most important challenges to valid causal inference—for example, confounding variables, uncovering causal order, or specifying the correct set of included variables. Nor does the algorithm itself provide guidance on which variables to treat as causes, contextual factors moderating the effects of causes, or mediators—i.e., mechanisms transmitting the effect of the causes.. In sum, the conditional associations

uncovered by CART do not themselves provide a causal inference. Additional information must be introduced.

*New Approaches to Interactions.* Quantitative work on interactions has moved the discussion well beyond STCM, and also beyond the traditional approach in quantitative research of representing interactions with multiplicative terms in regression analysis. Unlike STCM, this work accommodates diverse levels of measurement; and in contrast to some work in a regression framework, it favors simple data displays, arguing that findings are most productively presented in tables and graphs (for example, Kam and Franzese 2007: 26–27; 60–61). As with other standard approaches to interactions, these authors advocate close attention to contrasts in the effect of a given variable at different scores or levels of that variable. Such a view of interactions is inevitably precluded in STCM, as the dichotomization of all variables in the truth table eliminates the option of observing patterns of interaction across different levels of specific variables. Overall, in a period of the complexification of many methods, this approach to interactions has the merit of pointing toward simpler forms of analysis.

*Logistic Regression.* Variants of logistic regression—i.e., regression with a dichotomous dependent variable—have been discussed as alternative approaches to analytic tasks addressed by STCM.

Logistic regression is of course a standard approach in the toolkit of conventional quantitative methods, and like other such methods it has substantial drawbacks. Even as a means of describing relationships in a data set, logistic regression involves substantial complexity. And to interpret fitted values—i.e., the logistic distribution function, evaluated as a linear combination of regressors and coefficients—as predicted “probabilities” requires commitment to a host of modeling assumptions that may be tenuous in typical applications (Freedman 2009, Chapter 7).

However, these tools are noted briefly here because they are used by scholars who are part of the debate on STCM. Further, even for the researcher who has grave doubts about this modeling technique and does not wish to estimate the equations, the analytic setup may nonetheless provide a useful way of thinking about some issues raised by STCM. Logistic regression is therefore briefly noted here—but in the framework of great caution about its contribution.

One application of logistic regression could be to address the ex-post inevitability thesis, discussed above. For any given case, the real world yields only one outcome on the dependent variable, and inferring the explanation for that case has to rely on diverse forms of evidence, including—but not limited to—the particular outcome. Logit analysis opens the possibility of using a larger N to move beyond this limitation and, at least in principle, derive estimated probabilities for specific cases. Thus, the researcher can not only reason about “what actually happened,” but can also estimate its likelihood. Again, the credibility of that likelihood depends on the validity of the underlying model—but in contrast to STCM, such techniques have the heuristic value of bringing into focus the contrast between what actually happened and its likelihood of occurring.

### **6c. Assessing Mechanisms**

Understanding the mechanisms through which a cause affects an outcome is critical yet difficult. One reason mechanisms matter has to do with variation in context. Thus, if one understands the mechanism that produced the effect in one context, one can reason about whether the mechanism is likely to be operative in another. Mechanisms are therefore critical for assessing external validity (Cartwright and Hardie 2012). Yet, one cannot readily infer mechanisms through analysis of covariances in data. The best recent research on “mediation analysis”—unlike work on STCM—makes

clear the strong assumptions necessary to infer the effect of mediators or mechanisms.<sup>28</sup> Unfortunately, these assumptions are usually too strong for mediation analysis to be helpful in practice.

In light of this limitation, there appear to be two defensible approaches to inferring mechanisms. One is through design, by varying the independent variables to allow for implicit inferences about what the operative mechanism is. For example, if a given cause bundles two different potentially operative components, one could design a study that unbundles this causal factor into component parts (e.g. through factorial design) and see if only one or the other parts has an effect. A recent study by Bold et al. (2013) replicates previous studies showing a positive effect of contract teachers on educational attainment in Kenya. These authors unbundle the treatment by varying whether teachers are hired through NGOs (as in previous studies), or through the Kenyan educational bureaucracy. It turns out that hiring additional teachers only improves outcomes when NGOs do the hiring—perhaps because patronage politics or other factors drive hiring in the established bureaucracy. Thus, the results shed light on the conditions under which this intervention boosts educational attainment.

The other useful approach is to use causal-process observations (Collier, Brady, and Seawright 2010) in conjunction with strong theory. Information on causal process is tremendously useful for assessing the plausibility of different mechanisms that may link cause to effect—perhaps even more than data analysis which tries to assess how effects may vary at different levels of an independent variable. Despite STCM's emphasis on case knowledge, this kind of process-based information tends to drop out

---

<sup>28</sup> Work on STCM also does not typically distinguish between moderators (those variables that are realized before a notional treatment took place, so one can say whether the effect of X depends on the level of the moderating variable) and mediators or mechanisms, which are post-treatment and are plausibly an effect of the treatment itself.



of the picture when truth tables and other algorithms are applied to data from a moderate to large number of cases.

To conclude, the argument here is that making causal claims about interactions, contextual effects, and mechanisms is very difficult. The best new work on causal inference recognizes these limitations, and this leads to modesty in ambitions and claims. In contrast, recent work on STCM appears much less cautious. Again, STCM is motivated by admirable goals, but the capacity of the method to deliver on these goals is limited. The alternative options discussed in this section point to better avenues for achieving these goals—though these too face important challenges that should be emphasized in any application.

## **7. Toward a Conclusion**

This article is a work in progress, and the conclusion has not been mapped out. For now, three concluding points may be made.

1. The effort to draw the attention of researchers and methodologists to concepts, context, complexity, and cases is highly commendable.

2. The set-theoretic approach has been counterproductive. It slices political and social reality in a way that is not useful and obscures too much. Procedures for conceptualization, measurement, and data analysis are too often problematic and routinely fail to yield findings of interest to many scholars. Efforts to strengthen causal inference with process tracing appear to be undermined by building into this method the premise that the findings inherently involve relations of necessity and sufficiency.

3. As an alternative, some better options were explored in Part 7. These are presented in a spirit of caution about what is claimed for each. For many of them we have sounded a note of caution, twice saying—for example—that the method might

best be used as a heuristic, As Brady, Collier, and Seawright (2010: 22) put it, “both qualitative and quantitative research are hard to do well.” There are no panaceas, and comprehensive systems too often disappoint. What is needed is circumspection in selecting methods, along with resistance to the decades-long tradition in political science of an endless quest for the latest “chique technique.” Plus tenacity, rigor, and shoe leather.<sup>29</sup>

---

<sup>29</sup> See Freedman 1991.

## Bibliography

- Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory* 6(2): 169–86.
- Achen, Christopher H. 1977. "Measuring Representation: Perils of the Correlation Coefficient." *American Journal of Political Science* 21(4): 805.
- Achen, Christopher H. 1982. *Interpreting and Using Regression*. Beverly Hills, CA: Sage Publications.
- Achen, Christopher H. 1983. "Toward Theories of Data: The State of Political Methodology." *Political Science: The State of the Discipline*: 69–93.
- Achen, Christopher H. 1992. "Social Psychology, Demographic Variables, and Linear Regression: Breaking the Iron Triangle in Voting Research." *Political Behavior* 14(3): 195–211.
- Achen, Christopher H. 2002. "Toward a New Political Methodology: Microfoundations and ART." *Annual Review of Political Science* 5: 423–50.
- Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3): 529–546.
- Ahmed, Amel, and Rudra Sil. 2012. "When Multi-Method Research Subverts Methodological Pluralism— or, Why We Still Need Single-Method Research." *Perspectives on Politics* 10(04): 935–53.
- Alker, Hayward R. 1973. "On Political Capabilities in a Schedule Sense: Measuring Power, Integration, and Development." In *Mathematical Approaches to Politics*, eds. Hayward R. Alker, Karl W. Deutsch, and Antoine H. Stoetzel. Amsterdam: Elsevier, 307–74.
- Avdagic, Sabina. 2010. "When Are Concerted Reforms Feasible? Explaining the Emergence of Social Pacts in Western Europe." *Comparative Political Studies* 43(5): 628–57.
- Beach, Derek, and Rasmus Brun Pedersen. 2013. *Process-Tracing Methods Foundations and Guidelines*. Ann Arbor, MI: University of Michigan Press.
- Berg-Schlosser, Dirk. 2008. "Determinants of Democratic Successes and Failures in Africa." *European Journal of Political Research* 47(3): 269–306.
- Bennett, Andrew. 2010. "Process Tracing and Causal Inference." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Lanham, MD: Rowman and Littlefield, 207–19.
- Bennett, Andrew. 2014. "Process Tracing with Bayes: Moving beyond the Criteria of Necessity and Sufficiency." *Qualitative and Multi-Method Research* 12(1): 46–51.
- Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2013. "Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education." Center on Global Development, Working Paper 321.
- Bollen, Kenneth A., and Robert W. Jackman. 1989. "Democracy, Stability, and Dichotomies." *American Sociological Review* 54(4): 612.
- Brady, Henry E. 2004. "Data-Set Observations versus Causal-Process Observations: The 2000 U.S. Presidential Election." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, eds. Henry E. Brady and David Collier. Lanham, MD: Rowman and Littlefield, 267–71.
- Brady, Henry E. 2008. "Causation and Explanation in Social Science." In *Oxford Handbook of Political Methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. New York: Oxford University Press.
- Brady, Henry E., and David Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman and Littlefield.
- Brady, Henry E., David Collier, and Jason Seawright. 2010. "Refocusing the Discussion of Methodology." In Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman and Littlefield, 15–31.
- Braumoeller, Bear F. 2003. "Causal Complexity and the Study of Politics." *Political Analysis* 11(3): 209–33.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45: 5–32.
- Cartwright, Nancy and Jeremie Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford University Press.
- Cebotari, Victor, and Maarten P. Vink. 2013. "A Configurational Analysis of Ethnic Protest in Europe." *International Journal of Comparative Sociology* 54(4): 298–324.
- Collier, David. 2011. "Understanding Process Tracing." *PS: Political Science and Politics* 44(4): 823–30.
- Collier, D. 2014. "Comment: QCA Should Set Aside the Algorithms." *Sociological Methodology* 44(1): 122–26.
- Collier, David, Henry E. Brady, and Jason Seawright. 2010. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology." In Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman and Littlefield, 161–199.

- Collier, David, Jody LaPorte, and Jason Seawright. 2012. "Putting Typologies to Work: Concept Formation, Measurement, and Analytic Rigor." *Political Research Quarterly* 65(1): 217–32.
- Collier, Ruth Berins and David Collier. 1991. *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. Princeton, NJ: Princeton University Press.
- Collier, Ruth Berins. 1982. *Regimes in Tropical Africa: Changing Forms of Supremacy, 1945-1975*. Berkeley: University of California Press.
- Cronqvist, Lasse, and Dirk Berg-Schlosser. 2009. "4 Multi-Value QCA (mvQCA)." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, eds. Benoît Rihoux and Charles C Ragin. Thousand Oaks, CA: Sage Publications.
- Dahl, Robert A. 1971. *Polyarchy: Participation and Opposition*. New Haven: Yale University Press.
- Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge: Cambridge University Press.
- Dunning, Thad. 2013. "Contributions of Fuzzy-Set/Qualitative Comparative Analysis: Some Questions and Misgivings." Paper presented at the annual American Political Science Association in Chicago, Illinois. August 31–September 2.
- Elman, Colin. 2013. "Duck-Rabbits in Social Analysis: A Tale of Two Cultures." *Comparative Political Studies* 46(2): 266–77.
- Freedman, David A. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology* 21: 291–313.
- Freedman, David A. 2008. "Randomization Does Not Justify Logistic Regression." *Statistical Science* 23(2): 237–49.
- Freedman, David A. 2009. *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press.
- Gerring, John. 2008. "Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques." In *Oxford Handbook of Political Methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford: Oxford University Press, 645–84.
- Goertz, Gary. 2003. "Cause, Correlation, and Necessary Conditions." In *Necessary Conditions: Theory, Methodology, and Applications*, eds. Gary Goertz and Harvey Starr. Lanham, Maryland: Rowman and Littlefield, 47–64.
- Goertz, Gary, and James Mahoney. 2012a. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.
- Goertz, Gary, and James Mahoney. 2012b. "Concepts and Measurement: Ontology and Epistemology." *Social Science Information* 51(2): 205–16.
- Goodman, Leo A., and William H. Kruskal. 1954. "Measures of Association for Cross Classifications." *Journal of the American Statistical Association* 49(268): 732–64.
- Granato, Jim, and Frank Scioli. 2004. "Puzzles, Proverbs, and Omega Matrices: The Scientific and Social Significance of Empirical Implications of Theoretical Models (EITM)." *Perspectives on Politics* 2(02): 313–23.
- Hartmann, Christof, and Jörg Kemmerzell. 2010. "Understanding Variations in Party Bans in Africa." *Democratization* 17(4): 642–65.
- Hill, Daniel W., and Zachary M. Jones. 2014. "An Empirical Evaluation of Explanations for State Repression." *American Political Science Review* 108(03): 661–87.
- Hug, Simon. 2013. "Qualitative Comparative Analysis: How Inductive Use and Measurement Error Lead to Problematic Inference." *Political Analysis* 21(2): 252–65.
- Humphreys, Macartan, and Alan Jacobs. 2013. "Mixing Methods: A Bayesian Unification of Qualitative and Quantitative Approaches." Department of Political Science, Columbia University.
- Kam, Cindy D., and Robert J. Franzese. 2007. *Modeling and Interpreting: Interactive Hypotheses in Regression Analysis*. Ann Arbor, MI: University of Michigan Press.
- Kosko, Bart. 1993. *Fuzzy Thinking: The New Science of Fuzzy Logic*. New York: Hyperion.
- Krogslund, Christopher, Donghyun Danny Choi, and Matthias Poertner. 2015. "Fuzzy Sets on Shaky Ground: Parameter Sensitivity and Confirmation Bias in fsQCA." *Political Analysis*, forthcoming.
- Krogslund, Christopher, and Katherine Michel. 2014. "Can QCA Uncover Causal Relationships? An Assessment and Proposed Alternative." Paper Presented at the 2014 Annual Meeting of American Political Science Association, Chicago IL. August 31–September 2.
- Kurtz, Marcus. 2013. "The Promise and Perils of Fuzzy-set/Qualitative Comparative Analysis: Measurement Error and the Limits of Inference." Paper Presented at the 2013 Annual Meeting of the American Political Science Association, Chicago IL. August 31–September 2.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

- Lakoff, George. 2014. "Set Theory and Fuzzy Sets: Their Relationship to Natural Language. Interview with George Lakoff, Conducted by Roxanna Ramzipoor." *Qualitative and Multi-Method Research* 12(1): 9–13.
- Lakoff, George, and Mark Johnson. 1980. *Metaphors We Live by*. Chicago: University of Chicago Press.
- Lazarsfeld, Paul F. 1955. "Interpretation of Statistical Relations as a Research Operation." In *The Language of Social Research*, eds. Paul F. Lazarsfeld and Morris Rosenberg. New York: Free Press.
- Lieberman, Evan S. 2003. *Race and Regionalism in the Politics of Taxation in Brazil and South Africa*. Cambridge: Cambridge University Press.
- Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99(3): 435–52.
- Lieberson, Stanley. 1987. "Asymmetrical Forms of Causation." In *Making It Count: The Improvement of Social Research and Theory*, Berkeley and Los Angeles, CA: University of California Press.
- Linz, Juan J. 1975. "Totalitarian and Authoritarian Regimes." In *Handbook of Political Science: Macropolitical Theory v. 3*, eds. Greenstein, Fred I., and Nelson W. Polsby. Reading, MA: Addison-Wesley. 175–412.
- Locke, Richard M., and Kathleen Thelen. 1995. "Apples and Oranges: Contextualized Comparisons and the Study of Comparative Labor Politics." *Politics and Society* 23(3): 337–67.
- Locke, Richard M., and Kathleen Thelen. 1995. "Apples and Oranges Revisited: Contextualized Comparisons and the Study of Comparative Labor Politics." *Politics and Society* 23(3): 337–67.
- Lucas, Samuel R., and Alisa Szatrowski. 2014. "Qualitative Comparative Analysis in Critical Perspective." *Sociological Methodology* 44(1): 1–79.
- Mahoney, James. 2008. "Toward a Unified Theory of Causality." *Comparative Political Studies* 41(4-5): 412–36.
- Mahoney, James. 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62(1): 120–47.
- Mahoney, James. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods and Research* 41(4): 570–97.
- Mauldon, Jane, Jan Malvin, John Stiles, Nancy Nicosia and Eva Y. Seto. 2000. *The Impact of California's Cal-Learn Demonstration Project*. UC Data Archive and Technical Assistance, Survey Research Center, UC Berkeley. Retrieved from <http://escholarship.org/uc/item/2np332fc>. Viewed January 14, 2014.
- Pajunen, Kalle. 2008. "Institutions and Inflows of Foreign Direct Investment: A Fuzzy-Set Analysis." *Journal of International Business Studies* 39(4): 652–69.
- Przeworski, Adam, Michael E. Alvarez, Jose Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*. 1st edn. Cambridge: Cambridge University Press.
- Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience.
- Ragin, Charles C., Susan E. Mayer, and Kriss A. Drass. 1984. "Assessing Discrimination: A Boolean Approach." *American Sociological Review* 49(2): 221.
- 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: The Unity and Diversity of Method*. 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. "Comment Lucas and Szatrowski in Critical Perspective." *Sociological Methodology* 44: 80-94.
- Ragin, Charles C., and Lisa M. Amoroso. 2010. *Constructing Social Research: The Unity and Diversity of Method*. 2nd edn. Los Angeles, CA: Sage Publications.
- Ragin, Charles C., and Howard S. Becker. 1989. "How the Microcomputer Is Changing Our Analytic Habits." In *New Technology in Society: Practical Applications in Research and Work*, New Brunswick, N.J.: 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., Susan E. Mayer, and Kriss A. Drass. 1984. "Assessing Discrimination: A Boolean Approach." *American Sociological Review* 49(2): 221.

- Ragin, Charles C., and David Zaret. 1983. "Theory and Method in Comparative Research: Two Strategies." *Social Forces* 61(3): 731–54.
- Ridenour, Professor Carolyn S., and Professor Isadore Newman. 2008. *Mixed Methods Research: Exploring the Interactive Continuum*. 2nd edn. Carbondale: Southern Illinois University Press.
- Rihoux, Benoît, and Gisèle De Meur. 2009. "Crisp-Set Qualitative Comparative Analysis (csQCA)." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, Thousand Oaks, CA: Sage Publications, 33–68.
- Rihoux, Benoît, and Charles C. Ragin. 2009. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, CA: Sage Publications.
- 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): 220–30.
- Rohlfing, Ingo and Carsten Q. Schneider. "The Formalized Choice of Cases for Comparative Process Tracing Based on QCA Results." Paper Presented at the 2014 Annual Meeting of the American Political Science Association, Washington, DC, 2014.
- Safaei, Javad, and Hamid Beigy. 2007. "Quine-McCluskey Classification." In IEEE/ACS International Conference on Computer Systems and Applications, 2007. AICCSA '07, 404–11.
- Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64(4): 1033–53.
- Sartori, Giovanni. 2014. "Logic and Set Theory: A Note of Dissent." *Qualitative and Multi-Method Research* 12 (1): 14–15.
- Schrodt, Philip A. 2002. "Review of Fuzzy-Set Social Science by Charles Ragin (Chicago: University of Chicago Press)." *American Political Science Review* 96(2): 452–453.
- Schneider, Carsten Q., and Ingo Rohlfing. 2013. "Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research." *Sociological Methods Research* 42(4): 559–97.
- Schneider, Carsten Q., and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge: Cambridge University Press.
- Schneider, Carsten Q., and Claudius Wagemann. 2013. "Are We All Set?" *Qualitative and Multi-Method Research* 11(1): 5–8.
- Seawright, Jason. 2005. "Qualitative Comparative Analysis Vis-a-Vis Regression." *Studies in Comparative International Development* 40(1): 3–26.
- Seawright, Jason. 2013. "Warrantable and Unwarranted Methods: The Case of QCA." Paper Presented at the 2013 Annual Meeting of the American Political Science Association, Chicago IL.
- Seawright, Jason, and John Gerring. 2008. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61(2): 294–308.
- Sekhon, Jasjeet S. 2008. "The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods." In *Oxford Handbook of Political Methodology*, eds. Janet Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford: Oxford University Press, 271–99.
- Skocpol, Theda. 1995. *Protecting Soldiers and Mothers*. Harvard University Press.
- 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): 331–38.
- Tannenwald, Nina. 1999. "The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use." *International Organization* 53(3): 433–68.
- Tanner, Sean. 2014. "Evaluating QCA: A Poor Match for Public Policy Research." *Qualitative & Multi-Method Research* 12(1): Forthcoming.
- 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.
- Thiem, Alrik, and Adrian Dusa. 2013. "When More Than Time Is of the Essence: Enhancing QCA with eQMC." Paper Presented at the 2013 Annual Meeting of the American Political Science Association, Chicago.
- Van Deth, Jan W. 1998/2013. *Comparative Politics: The Problem of Equivalence*. London: Routledge, Reissued in the ECPR Classics Series, European Consortium for Political Research Press.
- World Bank. 2014. *Social Gains in the Balance: A Fiscal Policy Challenge for Latin America and the Caribbean*. Washington, D.C.: World Bank Document 85162.
- Zadeh, Lotfi A. 1965. "Fuzzy Sets." *Information and Control* 8: 338–53.
- Zadeh, Lotfi A. 1971. "Similarity Relations and Fuzzy Orderings." *Information Sciences* 3(2): 177–200.
- Zaks, Sherry. 2013. "Quest for an Analytic Middle Ground: Assessing the Claims of QCA." Paper presented at the Annual Meeting of the American Political Science Association, Chicago.